

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:40:09 ; Search time 26.7893 Seconds
(without alignments)

646,490 Million cell updates/sec

Title: US-10-699-035a-2

Perfect score: 913

Sequence: 1 RGDLMFLDSSASVSHYFSS.....FVDDHLHIVQELRGSIID 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	265.5	29.1	1857	2 S31212	collagen alpha 1(X)
2	265.5	29.1	1888	2 S74476	collagen alpha 1(X)
3	265.5	29.1	3124	2 A40020	collagen alpha 1(X)
4	254.5	27.9	1747	2 A45974	collagen alpha 1(X)
5	247.5	27.1	493	2 A33809	cartilage matrix P
6	239.5	26.2	500	2 S65222	cartilage matrix P
7	233.5	25.6	496	2 A37979	collagen alpha 1(V)
8	227.5	24.9	2944	2 A54849	collagen alpha 1(V)
9	216.5	23.7	3137	2 A37797	collagen alpha 3(V)
10	214	23.4	741	2 T46488	hypothetical prote
11	207.5	22.7	3176	2 CGH3A	collagen alpha 3(V)
12	198.5	21.7	929	2 I51027	type XII collagen
13	198	21.7	272	2 A55348	integrin alpha-1 -
14	194	21.2	1151	2 A45226	integrin alpha-1 c
15	191.5	21.0	843	2 A40970	undulin 1 - human
16	184	20.2	1180	2 A35854	hypothetical prote
17	179	19.6	550	2 T23760	hypothetical prote
18	175	19.2	724	2 A48569	antigen Em100 - E1
19	175	19.2	1286	2 A88396	protein M01E10.2 (
20	168	18.4	3051	2 S42373	hypothetical prote
21	162.5	17.8	1153	1 RWH1B	cell surface glyco
22	162.5	17.8	1170	2 I45914	integrin alpha 2 s
23	159.5	17.5	1153	2 S00551	leukocyte surface
24	158.5	17.4	1178	2 S44142	VLA-2 protein homo
25	158	17.3	712	2 A45638	immunodominant mic
26	153	16.8	1163	2 I56126	lymphocyte function
27	152.5	16.7	427	2 G00039	von Willebrand fac
28	151.5	16.6	1163	1 RWH1C	cell surface glyco
29	145.5	15.9	1181	2 A33998	integrin alpha-2 c

30	144.5	15.8	2813	1 VWHU	von Willebrand fac
31	140	15.3	414	2 P50323	von Willebrand fac
32	139.5	15.3	191	2 I47230	VLA-2 protein - pi
33	133.5	14.6	1179	2 A53213	integrin alpha-E c
34	133	14.6	1170	2 S03308	cell surface glyco
35	113	12.4	640	2 T29784	hypothetical prote
36	111	12.0	13055	2 T16580	hypothetical prote
37	109.5	12.0	567	2 T28797	hypothetical prote
38	104	11.4	643	2 T19549	hypothetical prote
39	103.5	11.3	1028	1 CGH1A	collagen alpha 1(V)
40	99.5	10.9	475	2 T36921	hypothetical prote
41	99.5	10.9	1019	1 A32856	collagen alpha 1(V)
42	98.5	10.8	371	2 S32604	collagen alpha 2(V)
43	96.5	10.6	1025	2 S34839	collagen alpha 1(V)
44	96	10.5	918	2 S23377	collagen alpha 2(V)
45	96	10.5	1022	2 S04111	collagen alpha 2(V)

ALIGNMENTS

RESULT 1

S31212 collagen alpha 1(XIV) chain precursor, short form - chicken

C/Species: Gallus gallus (chicken)

C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 15-Sep-2003

C/Accession: S31212

R/Maeichli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.

Bur, J.; Biochem 212, 483-490, 1993

A/Title: Complete primary structure of chicken collagen XIV.

A/Reference number: S31211; MUID:93185668; PMID:8444186

A/Accession: S31212

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: mRNA

A/Residues: 1-1857 <WAB>

A/Cross-references: UNIPARC:UPI000017123C; EMBL:X70792; NID:G288674; PIDN:CA50063.1; PI

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

C/Genetics:

A/Gene: Col14A1

C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trinc

F.1-28/Domain: signal sequence #status predicted <SIG>

F.129-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>

F.156-110/Domain: fibronectin type III repeat homology <FN3A>

F.156-320/Domain: von Willebrand factor type A repeat homology <VWA1>

F.352-433/Domain: fibronectin type III repeat homology <FN3B>

F.442-525/Domain: fibronectin type III repeat homology <FN3C>

F.534-614/Domain: fibronectin type III repeat homology <FN3D>

F.623-707/Domain: fibronectin type III repeat homology <FN3E>

F.741-823/Domain: fibronectin type III repeat homology <FN3F>

F.832-914/Domain: fibronectin type III repeat homology <FN3G>

F.922-1009/Domain: fibronectin type III repeat homology <FN3H>

F.1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 29.1%; Score 265.5; DB 2; Length 1857;

Best Local Similarity 34.5%; Pred. No. 2.2e-15;

Matches 61; Conservative 33; Mismatches 82; Indels 1; Gaps 1;

QY	3	DLMFLDSSASVSHYFSSRVFVQGLVAPLPLGALRASLVHSGSPRYTEPFGGSS	62
DB	158	DIIVLVDSWISGRNFRFLVLENTLSAFNVGSEKTRVGLAQSGBPRTIMHINAYGT	217
QY	63	GKAQDAVRAASQKMGDTHTGILATYAAEQLEFAEASGARPGVKLVWTTDGSDDPYGP	122
DB	218	KDAVDAVRNLPYKGNLTGLATLYIENSFKPKAGARPGVSKIGILITDQSDVDYIP	277
QY	123	PMQELKDGVTIVYVSTGRNFLLESAASAPAEAGHLFV-DVDDHLHIVQELRCSI	178
DB	278	PAKNLRDAGIEHFAIGVNVADINEIKIASBDSPTHVNVADFNFMNSIVGELTRTV	334

RESULT 2
S78476 collagen alpha 1(XIV) chain precursor, long form - chicken

C/Species: Gallus gallus (chicken)
C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
C/Accession: S78476; S31211
R/Trueb, B.
Submitted to the EMBL Data Library, January 1993
A/Reference number: S78476
A/Accession: S78476
A/Molecule type: mRNA
A/Residues: 1-1888 <TRU>
A/Cross-references: UNIPROT:P32018; UNIPARC:UPI0000126D31; EMBL:X70793; NID:G288872; PID:R/Melchli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A/Title: Complete primary structure of chicken collagen XIV.
A/Reference number: S31211; MUID:93185668; PMID:8444186
A/Accession: S31211
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-416;1460-1811,1843-1888 <MA>
A/Cross-references: UNIPARC:UPI00001773F4; UNIPARC:UPI00001773F5; EMBL:X70793
C/Genetics:
A/Gene: Col14A1
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>
F:29-110/Domain: fibronectin type III repeat homology <FN3A>
F:156-330/Domain: von Willebrand factor type A repeat homology <VWA1>
F:352-433/Domain: fibronectin type III repeat homology <FN3B>
F:442-525/Domain: fibronectin type III repeat homology <FN3D>
F:534-614/Domain: fibronectin type III repeat homology <FN3E>
F:623-707/Domain: fibronectin type III repeat homology <FN3F>
F:741-823/Domain: fibronectin type III repeat homology <FN3G>
F:833-914/Domain: fibronectin type III repeat homology <FN3H>
F:922-1009/Domain: fibronectin type III repeat homology <FN3I>
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 29.1%; Score 265.5; DB 2; Length 1888;
Best Local Similarity 34.5%; Pred. No. 2.3e-15;
Matches 61; Conservative 33; Mismatches 82; Indels 1; Gaps 1;

Oy 3 DLMFLDSSASVSHFEFGRVREFVQGLVAPLPLGTGALRASIVNHVGSRPYEPFGGSS 62
Db 158 DIVILVDSWSIGRNFPLVLFLENLVSARFVSGSEKRVGLAQSGDPRLEMLNAYGT 217

Oy 63 GEAAQDAVRASQRMGDTHTGTALVYAKQQLFAEASGARPGVKLVWVTDGSSDPYGP 122
Db 218 KDAVLDVAVRNLPYKGNLTLTGALTYYIENSPKPPAGARPGVSKIGITLTDGKSDVDYIP 277

Oy 123 PMQELKDIQVTVFVSTGRGNLELSAASAPAEHGHPV-DVDDHLIIYQELRQSI 178
Db 278 PAKNLRDGIETFAIGVKNADINELKEIASPEPDSTHYVNADPFMNSIVEGLTTRTV 334

RESULT 3
A40020
collagen alpha 1(XII) chain precursor - chicken
N/Alternate names: fibrochimarin
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
R/Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obari, J. Cell Biol. 115, 209-221, 1991
A/Title: The complete primary structure of type XII collagen shows a chimeric molecule within region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.
A/Reference number: A40020; MUID:92011862; PMID:1918137
A/Accession: A40020
A/Molecule type: mRNA
A/Residues: 1-3124 <YAM>
A/Cross-references: UNIPROT:P13944; UNIPARC:UPI0000126D2D; GB:D00824; NID:9222810; PIDN:A/Note: in the authors' translation residues 1216-1219 are shown after residue 1235 and, R/Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989
A/Title: Type XII collagen. A large multidomain molecule with partial homology to type I
A/Reference number: A34485; MUID:930062079; PMID:2584192

A/Accession: A34485
A/Molecule type: mRNA
A/Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A/Cross-references: UNIPARC:UPI0000171233; EMBL:J05137; NID:9211284; PIDN:AAA48635.1; PI
A/Accession: B34485
A/Molecule type: protein
A/Residues: 2772-2792;2846-2873 <GOR2>
A/Cross-references: UNIPARC:UPI0000173C3F; UNIPARC:UPI0000173C40
R/Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A/Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c
A/Reference number: A28037; MUID:87317590; PMID:3476925
A/Accession: A28037
A/Molecule type: mRNA
A/Residues: 2960-2976, 'F', 2978-3074, 'AG' <GOR3>
A/Cross-references: UNIPARC:UPI00001712F4; EMBL:M17375; NID:9211649; PIDN:AAA48718.1; PI
A/Note: this sequence has been revised in reference A34485
R/Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A/Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A/Reference number: S23814; MUID:92362621; PMID:1323460
A/Accession: S23814
A/Molecule type: protein
A/Residues: 'X',1333,'Q',1335-1347;1914-1928;2504,'X',2506,'X',2508-2511,'X',2513-2517 <
A/Cross-references: UNIPARC:UPI0000173C41; UNIPARC:UPI0000173C42; UNIPARC:UPI0000173C43
R/Dublet, B.; van der Rest, M.
J. Biol. Chem. 262, 17724-17727, 1987
A/Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepain-
A/Reference number: S22254; MUID:88087065; PMID:33121603
A/Accession: S22254
A/Molecule type: protein
A/Residues: 2831-2833,'T',2834,'R',2836-2843;3002-3014 <DUB>
A/Cross-references: UNIPARC:UPI0000173C44; UNIPARC:UPI0000173C45
R/Trueb, J.; Trueb, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A/Title: The two splice variants of collagen XII share a common 5' end.
A/Reference number: S28811; MUID:93042014; PMID:1420368
A/Accession: S28811
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-24,1189-1257,'S',1259-1263,'E',1265-1280 <TRU>
A/Cross-references: UNIPARC:UPI0000173C46; EMBL:X67327
C/Genetics:
A/Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
C/Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfid
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F:24-114/Domain: II1A #status predicted <II1A>
F:124-105/Domain: fibronectin type III repeat homology <FN3A>
F:133-425/Domain: II1B #status predicted <II1B>
F:333-414/Domain: fibronectin type III repeat homology <FN3B>
F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
F:623-1178/Domain: II1C #status predicted <II1C>
F:630-711/Domain: fibronectin type III repeat homology <FN3C>
F:721-802/Domain: fibronectin type III repeat homology <FN3D>
F:812-895/Domain: fibronectin type III repeat homology <FN3E>
F:905-986/Domain: fibronectin type III repeat homology <FN3F>
F:995-1076/Domain: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>
F:1384-2295/Domain: II1D #status predicted <II1D>
F:1384-1465/Domain: fibronectin type III repeat homology <FN3I>
F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>
F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>
F:1655-1738/Domain: fibronectin type III repeat homology <FN3L>
F:1756-1838/Domain: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>
F:2119-2199/Domain: fibronectin type III repeat homology <FN3Q>
F:2207-2294/Domain: fibronectin type III repeat homology <FN3R>

[illegible]

Db 332 KKDIIKAAVKKKAAVNEKGMGTGALKLYLVDSFSFIANGARPEVPRKGVIFVTDGRSQDYITD 391

Qy 123 PMQELKDLGVTVFVIVSTGRGNFLELSAAASAPAEKHLFVDVD 165

Db 392 AAKRAKDLGFRMFVAVGVNAVEDELREIASPEPAEH-YFYTAD 433

RESULT 6

S66532

cartilage matrix protein precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004

A:Accession: S66522

R:Ascoli, A.; Hauser, N.; Studer, D.; Paulsson, M.; Hirtl, L.; Bosse, Z.

Eur. J. Biochem. 236, 970-977, 1996

A:Title: Cloning, sequencing and expression analysis of mouse cartilage matrix protein c

A:Reference number: S66522; MUID:96270751; PMID:8665920

A:Accession: S66522

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-500 <ASZ>

A:Cross-references: UNIPROT:P51942; UNIPARC:UPI0000029480; EMBL:U35035; NID:g1163178; PE

C:Genetics:

A:Gene: CMP

C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repeat

F:1-29/Domains: signal sequence #status predicted <SIG>

F:30-500/Product: cartilage matrix protein #status predicted <MAT>

F:43-210/Domains: von Willebrand factor type A repeat homology <WVA1>

F:231-266/Domains: EGF homology <EGF>

F:277-441/Domains: von Willebrand factor type A repeat homology <WMA2>

Query Match 26.2%; Score 239.5; DB 2; Length 500;

Best Local Similarity 34.8%; Pred. No. 1e-13;

Matches 57; Conservative 29; Mismatches 75; Indels 3; Gaps 1;

Qy 3 DLMFLDSSASVSHYEFGRVREFVQGLVAPLPLGNGALRASLVHVGSRPTEPFQGHSS 62

Db 45 DLVFPVDSRSRVRPEVEFKVFLSQVLESIDVGNATRVGLVNVASTVKEPFLRAHGS 104

Qy 63 GEAQDAPARASAOQMGDHTHTGLATVYAKEQLFAEASGAR--PGVPKVLVWVTDGSSDP 119

Db 105 KASLLQAVRRRQPLSTGMTGLALQFALTAKLSDEGGRASPDISKVLIVVTDGRQDS 164

Qy 120 VGPPMQELKDLGVTVFVIVSTGRGNFLELSAAASAPAEKHLFVD 163

Db 165 VRDVSERARASGIELFALGLGRVDKATLRQLASEPQDEHYDVE 208

RESULT 7

A37979

cartilage matrix protein precursor - human

C:Species: Homo sapiens (man)

C:Date: 12-Jul-1991 #sequence_revision 12-Jul-1991 #text_change 09-Jul-2004

A:Accession: A37979

R:Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; Sh

J. Biol. Chem. 265, 19624-19631, 1990

A:Title: Structure and chromosomal location of the human gene encoding cartilage matrix

A:Reference number: A37979; MUID:91060568; PMID:2246248

A:Accession: A37979

A:Molecule type: DNA

A:Residues: 1-496 <JEN>

A:Cross-references: UNIPROT:P21941; UNIPARC:UPI000004F1ED; GB:J05667

A:Accession: B37979

A:Molecule type: mRNA

A:Residues: 157-290, 'L', 292-496 <JEN>

A:Cross-references: UNIPARC:UPI000016A6E8; GB:M55683; GB:J05666; GB:J05667; NID:g180651,

C:Genetics:

A:Gene: GDB:CRTM

A:Cross-references: GDB:127280; OMIM:115437

A:Map position: 1p35-1p35

A:Introns: 32/1; 147/3; 222/1; 264/1; 403/1; 454/1; 481/1

C:Complex: homotrimer

C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe

C:Keywords: glycoprotein; homotrimer
F1:22/Domain: signal sequence #status predicted <Sig>
F1:23-496/Product: cartilage matrix protein #status predicted <MAT>
F1:39-206/Domain: von Willebrand factor type A repeat homology <VMA>
F1:227-262/Domain: EGF homology <EGF>
F1:227-437/Domain: von Willebrand factor type A repeat homology <VMA>
F1:76-344/Binding site: carbohydrate (Asn) (covalent) #status predicted
F1:221-238,234-247,249-262/Disulfide bonds: #status predicted

```
Query Match          25.6%; Score 233.5; DB 2; Length 496;
Best Local Similarity 35.0%; Pred. No. 3,5e-13;
Matches 62; Conservative 27; Mismatches 87; Indels 1; Gaps 1;
```

OY 3 DLMFLDSSASVSHPEFVSREFVQQLVAPPLPGTALASLVHVSGRPTTEPPFGHSS 62
DB 275 DLVFLIDSKSRPENFELVKRFISQIVDTLVSDKLAQVGLQVSSVRQEPFLGRFT 334
OY 63 GEAAADAVARASQRMGDFHTGALVYAEQQLFAEASGAPGVKVLVWVTDGSSDPVGP 122
DB 335 KKDTIAARNRNSYMERKMTGAAALKTLIDNSFVSSGAPGAKQGVLFVDGSSQDYIND 394
OY 123 PMQELKDLGVTVFIVSTGRNFLELSAASAP-AEKHLHFVDVDDHLIIQELRGS1 178
DB 395 AAKKAKDGFKNFAVGVAVEDELREIASEPVAEHYFTADPKTINQICKLQKKI 451

RESULT 8
A54849
collagen alpha 1(VII) chain precursor - human
M:Alternate names: procollagen alpha 1(VII) chain
C:Species: Homo sapiens (man)
C:Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #ext. change 09-Jul-2004
C:Accession: A54849; PH0844; S16316; I56328; A30256; I84686
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Oltro, J.
J. Biol. Chem. 269, 20256-20262, 1994
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha (VII)
A:Reference number: A54849; MUID:94327588; PMID:8051117
A:Accession: A54849
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2944 <CHR>
A:Molecule type: mRNA
A:Residues: 1-2944 <CHR>
A:Cross-references: UNIPROT:002388; UNIPARC:UPI000017A138; GB:U02870; NID:9987124; PIDN:1
R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A:Title: Molecular cloning and characterization of type VII collagen cDNA.
A:Reference number: PH0844; MUID:92231902; PMID:1567409
A:Accession: PH0844
A:Molecule type: mRNA
A:Residues: 'EFF',540-475,'RALSTSHSTLCKRAIRWPCNGSHHTRAACEPCNRPAHSRARAG',524-528,'C',5
A:Cross-references: UNIPARC:UPI00017A139; DDBU:DI1152; DDBU:DI1694; NID:9453698; PIDN:B
A:Experimental source: keratinocyte
A:Note: The authors translated the codon ACC for residues 394 and 397 as Tyr
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A:Reference number: S16316; MUID:91334380; PMID:1871109
A:Accession: S16316
A:Molecule type: mRNA
A:Residues: 815-892,'E',894-1439 <PAR>
A:Cross-references: UNIPARC:UPI000016A722; GB:M65158; GB:S49017; NID:gl80914; PIDN:AAA96
A:Experimental source: keratinocyte
R:Gammou, W.R.; Abornehy, M.L.; Padilla, K.M.; Prisayanah, P.S.; Cook, M.E.; Wright, J.;
J. Invest. Dermatol. 99, 691-696, 1992
A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot.
A:Reference number: I56328; MUID:93107742; PMID:1469284
A:Accession: I56328
A:Status: translated from GB/EMBL/DDBU
A:Molecule type: mRNA
A:Residues: 'EFF',372-517,'DV',520-540,'W',542-1255 <RES>
A:Cross-references: UNIPARC:UPI000016B3AC; GB:S51236; NID:9262308; PIDN:AAB24637.1; PID:5
R:Seitzler, J.L.; Elsen, A.Z.; Bauer, E.A.; Morris, N.P.; Glenville, R.W.; Burgeson, R.E.
J. Biol. Chem. 264, 3822-3826, 1989
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase

Db 157 DTAORLKGQGVKLFPAVGIRKNADEELKRVASQPTSDFFPFVN 199

```

RESULT 9
A:37797
collagen alpha 3 (VI) chain precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #ext_change 16-Aug-2004
C:Accession: A37797; A34270; A32674
R:Dioliana, R.; Bonaldo, P.; Colombatti, A.
A:Title: Multiple forms of chicken alpha3 (VI) collagen chain generated by alternative splicing
A:Reference number: A37797; PMID:91035630; PMID:1977751
A:Accession: A37797
A:Molecule type: mRNA
A:Residues: 1-253;312-321;434-453;504-518;635-655;704-717;832-853 <DOI>
A:Cross-references: UNIPROT:P15989; UNIPARC:UPI0000173C31; UNIPARC:UPI0000173C32; UNIPARC:M24282
R:Bonaldo, P.; Russo, V.; Buccicotti, F.; Dioliana, R.; Colombatti, A.
Biochemistry 29, 1245-1254, 1990
A:Title: Structural and functional features of the alpha3 chain indicate a bridging role
A:Reference number: A34270; PMID:2322559
A:Accession: A34270
A:Molecule type: mRNA
A:Residues: 224-2871 <BON>
A:Cross-references: UNIPARC:UPI0000173C38; GB:M24282
A:Note: the authors translated the codon TTC for residue 1916 as Leu and TTC for residue 1917 as Ile.
R:Bonaldo, P.; Colombatti, A.
J. Biol. Chem. 264, 20235-20239, 1989
A:Title: The carboxyl terminus of the chicken alpha3 chain of collagen VI is a unique motif
A:Reference number: A32674; PMID:9062147; PMID:2584214
A:Accession: A32674
A:Molecule type: mRNA
A:Residues: 2151-2199;2792-3137 <BO2>
A:Cross-references: UNIPARC:UPI0000173C39; UNIPARC:UPI0000173C3A; GB:M24282
C:Genetics:
A:Introns: 30/1; 236/1; 437/1; 638/1; 838/1
C:Superfamily: collagen VI
C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; extracellular matrix
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-3137/Product: collagen alpha 3 (VI) chain #status predicted <MAT>
F:32-2042/Domain: non-collagenous #status predicted <MNC>
F:326-2042/Domain: von Willebrand factor type A repeat homology <VM01>
F:326-202/Domain: von Willebrand factor type A repeat homology <VM02>
F:329-404/Domain: von Willebrand factor type A repeat homology <VM03>
F:442-607/Domain: von Willebrand factor type A repeat homology <VM03>
F:642-807/Domain: von Willebrand factor type A repeat homology <VM04>
F:840-1004/Domain: von Willebrand factor type A repeat homology <VM05>
F:1033-1197/Domain: von Willebrand factor type A repeat homology <VM06>
F:1137-1400/Domain: von Willebrand factor type A repeat homology <VM07>
F:1439-1604/Domain: von Willebrand factor type A repeat homology <VM08>
F:1639-1804/Domain: von Willebrand factor type A repeat homology <VM09>
F:1838-2010/Domain: von Willebrand factor type A repeat homology <VM10>
F:2043-2378/Domain: collagenous #status predicted <COL>
F:2045-2047/Region: cell attachment (R-G-D) motif
F:2153-2153/Region: cell attachment (R-G-D) motif
F:2153-2155/Region: cell attachment (R-G-D) motif
F:2153-2161/Region: cell attachment (R-G-D) motif
F:2379-3137/Region: non-collagenous #status predicted <CNC>
F:2405-2577/Domain: von Willebrand factor type A repeat homology <VM11>
F:2623-2806/Domain: von Willebrand factor type A repeat homology <VM12>
F:2803-2846/Domain: platelet glycoprotein Ib-like #status predicted <GPI>
F:2954-3039/Domain: fibronectin type III repeat homology <FN3>
F:3072-3122/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F:201,2084,2436,2563,2581,2683,2867,2920,3003/Binding site: carbohydrate (Asn) (covalent)
Query Match 23.7% Score 216.5; DB 2; Length 3137;
Best Local Similarity 33.5%; Pred. No. 1e-10;
Matches 60; Conservative 26; Mismatches 90; Indels 3; Gaps 2;
3 DLMEFLDSSASVSHYFSRRVFEVQGLVAPLPYGTGALRALSLVHVGSRPYTEPFPGHSS 62
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
38 DILFLVSSSIGSGEHQQLVREFLVDVVKALDIVGNDPFRALVQFSNGPHTPEQLNTYPS 97
```

A.Molecule type: protein
A.Residues: 574-586:965-973, 'X', 975-976:1306-1325:1361-1377:1381-1401:1473-1506, 'X', 1508-1962, 'X', 1964-1965:2018-2037:22374-2410:2445-2459:2466-2469, 'X', 2471-2474:2504-2509, 'X', A.Cross-references: UNIPARC:UPI00000173C15; UNIPARC:UPI00000173C16; UNIPARC:UPI00000173C17; C1C; UNIPARC:UPI0000173C1D; UNIPARC:UPI0000173C1E; UNIPARC:UPI00000173C1F; UNIPARC:UPI00010000173C25; UNIPARC:UPI0000173C26; UNIPARC:UPI00000173C27 R.Zanussi, S.; Dollana, R.; Segat, D.; Bonaldi, P.; Colombari, A.
J. Biol. Chem. 267, 24082-24089, 1992
A.Title: The human type VI collagen gene, mRNA and protein variants of the alpha3 chain
A.Reference number: S28776; PMID:93054780; PMID:13359440
A.Accession: A57083
A.Molecule type: DNA
A.Residues: 310-328 <ZAN>
A.Cross-references: UNIPARC:UPI0000173C28
A.Accession: S28776
A.Molecule type: mRNA
A.Residues: 32-126, 'AK', 129-136, 'L', 138-236 <ZAZ>
A.Cross-references: UNIPARC:UPI0000173C28; GB:S49942; PIDN:AA824261.1, PID:: A.Chu, M.L.; Mann, K.; Deutzmann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.; Bur, J. Biochem. 168, 309-317, 1987
A.Title: Characterization of three constituent chains of collagen type VI by peptide seq
A.Reference number: S00126; PMID:88029444; PMID:3665927
A.Accession: S00245
A.Molecule type: mRNA, protein
A.Residues: 2024-2046:2059-2156, 'R', 2203-2208, 'X', 2210-2211, 'X', 2213-2227:2228-2251:2314
A.Cross-references: UNIPARC:UPI0000016A705; UNIPARC:UPI00000173C29; UNIPARC:UPI00000173C2A;

A>Note: The mRNA portion of the sequence corresponds to residues 2092-2157
R.Chnu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R.
J. Biol. Chem. 263, 18601-18606, 1988
A>Title: Amino acid sequence of the triple-helical domain of human collagen type VI.
A.Reference number: A31952; MWID:89066644; PMID:3198591
A.Accession: C31952
A:Molecule type: mRNA
A.Residues: 2038-2373 <CH4>
A:Cross-references: UNIPARC:UPI0000173C2E; GB:J04211; GB:M20778
A>Note: parts of this sequence were determined by protein sequencing
R.Well, D.; Mattei, M.G.; Passage, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deutzmann, J.; Hum, Genet. 42, 435-445, 1988
A>Title: Cloning and chromosomal localization of human genes encoding the three chains o
A.Reference number: A29848; MWID:88161046; PMID:3348212
A.Accession: C29848
A:Molecule type: mRNA
A.Residues: 2092-2151 <WE1>
A:Cross-references: UNIPARC:UPI000016A720; GB:M27449; NID:g291919; PIDN:AAA52057.1; PID:
A>Note: part of this sequence was determined by protein sequencing

A:Title: Further characterization of the three polypeptide chains of bovine and human hA:Reference number: S26506; MUID:83209648; PMID:6852033
A:Accession: S26510
A:Molecule type: protein
A:Residues: 'SAIAGVAGVG' <JAN>
A:Cross-references: UNIPARC:UPI00000173C2F
A:Note: this sequence cannot be reliably placed and probably represents the results from R.Mayer, U.; Poeschl, E.; Nisch, R.; Specks, U.; Pan, T.C.; Chu, M.L.; Timpl, R. Eur. J. Biochem. 225, 573-580, 1994
A:Title: Recombinant expression and properties of the Kunitz-type protease-inhibitor modA:Reference number: S48709; MUID:95045506; PMID:7525281
A:Accession: S48709
A:Molecule type: mRNA
A:Residues: 'MRMIFFLCLAGRLAA', 3102-3176 <MAY>
A:Cross-references: UNIPARC:UPI0000173C30
A:Note: engineered sequence to allow isolation of the Kunitz-type proteinase inhibitor hA:Rnoux, B.; Merigau, K.; Saludjian, P.; Norris, F.; Norris, K.; Bjorn, S.; Olsen, O.; submitted to the Brookhaven Protein Data Bank, August 1994
A:Reference number: A52812; PDB:1KNT
A:Contents: annotation; X-ray crystallography, 1.6 angstroms, residues 3106-3160
A:Note: engineered sequence expressed in *Saccharomyces cerevisiae* strain mt-663
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C) are 5-hydroxylated and subsequently O-glycosylated.
C:Comment: The fibronectin type III repeat hydrophobic domain may be released during processing.
C:Genetics:

A:Gene: GDB:COL6A3
A:Cross-references: GDB:119066; OMIM:120250
A:Map position: 2q37.3-2q37.3
C:Complex: type VI collagen is a heterotrimer of one alpha 1(VI) chain (see PIR:CGH1A),
associations among trimer amino- and carboxyl-terminal domains (with disulfide bonds).
C:Function:
A:Description: structural component of extracellular tissue microfibrils associated with
C:Superfamily: collagen VI
C:Keywords: alternative splicing; blocked amino end; cell binding; coiled coil; extracel
F:1-25/Domain: signal sequence #status predicted <Sig>
F:26-3176/Product: collagen alpha 3(VI) chain #status predicted <AMT>
F:26-2037/Domain: amino-terminal nonhelical #status predicted <AMT>
F:26-30, 237-3176/Product: collagen alpha 3(VI) chain, splice form A9/N10(-) #status pred
F:337-403/Domain: von Willebrand factor type A repeat homology <VM01>
F:240-405/Domain: von Willebrand factor type A repeat homology <VM02>
F:1443-608/Domain: von Willebrand factor type A repeat homology <VM03>
F:637-802/Domain: von Willebrand factor type A repeat homology <VM04>
F:835-999/Domain: von Willebrand factor type A repeat homology <VM05>
F:1027-1191/Domain: von Willebrand factor type A repeat homology <VM06>
F:1231-1394/Domain: von Willebrand factor type A repeat homology <VM07>
F:1434-1599/Domain: von Willebrand factor type A repeat homology <VM08>
F:1637-1802/Domain: von Willebrand factor type A repeat homology <VM09>
F:1836-2005/Domain: von Willebrand factor type A repeat homology <VM10>
F:2038-2373/Region: interrupted helical
F:2040-2042/Region: cell attachment (R-G-D) motif
F:2136-2138/Region: cell attachment (R-G-D) motif
F:2146-2150/Region: cell attachment (R-G-D) motif
F:2154-2156/Region: cell attachment (R-G-D) motif
F:2370-2372/Region: cell attachment (R-G-D) motif
F:2374-3176/Domain: carboxyl-terminal nonhelical #status predicted <CMH>
F:2400-2571/Domain: von Willebrand factor type A repeat homology <VM11>
F:2610-2800/Domain: von Willebrand factor type A repeat homology <VM12>
F:2865-2986/Region: von Willebrand factor type A repeat homology <VM13>
F:2987-3072/Domain: fibronectin type III repeat homology <FN3>
F:3111-3161/Domain: animal knittz-type proteinase inhibitor homology <BPI>
F:326/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:108,116,202,251,2079,2331,2558,2677,2861,3036/Binding site: carbohydrate (Asn) (cova
F:2087/Disulfide bonds: interchain #status predicted
F:2100,2206,2239,2316,2319/Modified site: 4-hydroxyproline (Pro) #status experimental
F:2103,2209,2212,2332,2337/Binding site: carbohydrate (Lys) (covalent) #status experimen
F:3111,3161,3120-3144,3136-3157/Disulfide bonds: #status predicted
Query Match 22.7%; Score 207.5; DB 2; Length 3176;
Best Local Similarity 31.4%; Pred. No. 6.5e-10;
Matches 55; Conservative 28; Mismatches 89; Indels 3; Gaps 2;
QY 3 DLMFLDSSASVSHYEFRRVREVGOLVAPPLGTGALRASLVHSGRPYTFPPGQHS 62
DB 39 DIIFLDSSWITGEHFOVLRREFLYDVVKSILAVGENDFFHAFVQFNGNPHTEFLNTYRT 98
QY 63 GEAADAVRASAQRMGDTHGTALVYAKQLFAEASGAR--PGVPKVLVWMTDGGSSDPV 120
DB 99 KQEVLSHINMYIGTGTOTGKLEIYMOHSLTKAASGAPGVPIVVLTDGSHKDEL 158
QY 121 GPPMOELKDGLVTVIVSTGRGNFLELSAASAPAEKHL-HFVDDHLIIYQEL 174
DB 159 ALPSAELKQADVNFALIVEADDEGALKEIASPELPMHMFENFSLSDIYGNL 213
RESULT 12
151027
type XII collagen alpha-1 chain - eastern newt (fragment)
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: I51027
R:Wei, Y.; Yang, E.-V.; Klatte, K.P.; Taseava, R.A.
Dev. Biol. 168, 503-513, 1995
A:Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII collag
A:Reference number: I51027; MUID:95246925; PMID:7729585
A:Accession: I51027
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 1-929 <WE1>
A:Cross-references: UNIPROT:Q91145; UNIPARC:UPI000126D2E; EMBL:U19494; NID:G632647; PIR
F:155-236/Domain: fibronectin type III repeat homology <3FR>
F:631-795/Domain: von Willebrand factor type A repeat homology <VMA3>
Query Match 21.7%; Score 198.5; DB 2; Length 929;
Best Local Similarity 27.9%; Pred. No. 9.9e-10;
Matches 50; Conservative 37; Mismatches 91; Indels 1; Gaps 1;
QY 1 RDDLFLDSSASVSHYEFRRVREVGOLVAPPLGTGALRASLVHSGRPYTFPPGQHS 60
DB 631 QNDIYLVLDGSSISGRPPFKIVRNFSRVVEYFDIGSDRQIAVQYSGDPTEWQLNTH 690
QY 61 SGEAADA VRASAQRMGDTHGTALVYAKQLFAEASGARPGVPIVWMTDGGSSDPV 120
DB 691 KTKKSLMDVAANLPYKGGNTNTGSALEKLENNFRPGVMREKARKIAILTDGKSQDDI 750
QY 121 GPPMOELKDGLVTVIVSTGRGNFLELSAASAPAEKHLHFV-DVDDHLIIYQELRGS 178
DB 751 VAPSRKRVADGEIYAVGIRKADENELKEIASDPDELVMYVAVDFSLTNTVNDLTENV 809
RESULT 13
A55348
Integrin alpha-1 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
C:Accession: A55348
R:Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.
J. Biol. Chem. 269, 22811-22816, 1994
A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1.
A:Reference number: A55348; MUID:94357930; PMID:7521332
A:Accession: A55348
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-272 <KER>
A:Cross-references: UNIPROT:Q42094; UNIPARC:UPI00017A14E; GB:U10114
F:155-230/Domain: von Willebrand factor type A repeat homology <VMA2>
Query Match 21.7%; Score 198; DB 2; Length 272;
Best Local Similarity 34.2%; Pred. No. 2.6e-10;
Matches 65; Conservative 26; Mismatches 75; Indels 24; Gaps 8;
QY 3 DLMFLDSSASVSHYEFRRVREVGOLVAPPLGTGALRASLVHSGRPYTFPPGQHS 62
DB 57 DIVYLDSSNST--YPMSSVTAFLNSILRNMDDIGQGOVGIVYQGVVHEFYINTYST 114
QY 63 GEAADAVRASAQRMG-DTHGTALVYAKQLFAEASGARPGVPIVWMTDGGSSDPV 121
DB 115 TEEVMDALRIQRGCTGTMTALGIDTAREEAFTEAHGARCGVQKVMYIVTDGESHD--N 172
QY 122 PPMOEL----KDLGTVIV-----STGRGNF-----ELSAASAPAEKHLHFVDDV 166
DB 173 YRLQEVINDKCEBENIQRFPAIALIGSYSKGNLSTKFEVEIKSIASKPTKK--HFNVSDE 230
QY 167 --LHIIYQEL 174
DB 231 LALVTVIV 240
RESULT 14
A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Briesewitz, R.; Bpstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:44:15 ; Search time 0.959464 Seconds
(without alignments)
1468.904 Million cell updates/sec

Title: US-10-699-035A-1

Perfect score: 1049
Sequence: 1 ggggacctgattctctcgc.....tgaagggtccattctcgcg 537

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 97014 segs, 13122536 residues

Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/abs/ABSGWEE spool/US10699035/runat_13022006_062453_25634/2pp_query.fasta_1
-DB=Published Applications_AA_New -QPMF=fastan -SURFIX=fapbn -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blousum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abs92p
-USER=US10699035 @CGN 1.1.10 @runat_13022006_062453_25634 -NCP1=6 -ICPU=3
-NO MAP -NEC SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA_New.*
1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US05_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	902	86.0	445	6	US-10-453-372-2
2	261.5	24.9	3063	7	US-11-186-284-26
3	257.5	23.5	517	7	US-11-169-041-160
4	246.5	23.5	915	6	US-10-131-826A-294
5	246.5	23.5	956	7	US-11-113-424-39
6	197	18.8	214	7	US-11-192-449-6
7	197	18.8	214	7	US-11-192-449-9
8	188	17.9	214	7	US-11-192-449-5
9	183	17.4	678	6	US-10-063-703-34

10	183	17.4	678	7	US-11-102-240-34	Sequence 34, Appli
11	166.5	15.9	1152	7	US-11-080-026-4	Sequence 4, Appli
12	162	15.4	709	6	US-10-453-372-186	Sequence 186, Appli
13	161	15.3	709	6	US-10-453-372-180	Sequence 180, Appli
14	161	15.3	3568	6	US-10-453-372-194	Sequence 194, Appli
15	161	15.3	3570	6	US-10-453-372-178	Sequence 178, Appli
16	161	15.3	3570	6	US-10-453-372-196	Sequence 196, Appli
17	161	15.3	3570	6	US-10-453-372-198	Sequence 198, Appli
18	161	15.3	3570	6	US-10-453-372-200	Sequence 200, Appli
19	161	15.3	3570	6	US-10-453-372-202	Sequence 202, Appli
20	161	15.3	3570	6	US-10-453-372-204	Sequence 204, Appli
21	161	15.3	3570	6	US-10-453-372-206	Sequence 206, Appli
22	160	15.3	709	6	US-10-453-372-182	Sequence 182, Appli
23	155	14.8	709	6	US-10-453-372-184	Sequence 184, Appli
24	150.5	14.3	1167	6	US-10-601-368-18	Sequence 18, Appli
25	150	14.3	1161	6	US-10-601-368-24	Sequence 22, Appli
26	150	14.3	1166	6	US-10-601-368-22	Sequence 24, Appli
27	150	14.3	1188	6	US-10-601-368-21	Sequence 21, Appli
28	149.5	14.3	182	6	US-10-601-368-25	Sequence 25, Appli
29	145.5	13.9	1147	6	US-10-453-372-4	Sequence 4, Appli
30	145	13.8	1141	6	US-10-601-368-6	Sequence 6, Appli
31	145	13.8	1166	6	US-10-601-368-4	Sequence 4, Appli
32	145	13.8	1188	6	US-10-601-368-3	Sequence 3, Appli
33	145	13.8	1188	7	US-11-000-463-338	Sequence 338, Appli
34	145	13.8	1188	7	US-11-000-463-810	Sequence 810, Appli
35	144.5	13.8	182	6	US-10-601-368-7	Sequence 7, Appli
36	144.5	13.8	2764	6	US-10-995-561-691	Sequence 691, Appli
37	144.5	13.8	2813	6	US-10-995-561-688	Sequence 688, Appli
38	144.5	13.8	2919	6	US-10-821-234-1133	Sequence 1133, Appli
39	134.5	12.8	1179	7	US-11-097-125-1	Sequence 1, Appli
40	134.5	12.8	1196	6	US-10-995-561-921	Sequence 921, Appli
41	134	12.8	184	6	US-10-665-658-8	Sequence 8, Appli
42	133	12.7	184	6	US-10-665-658-7	Sequence 7, Appli
43	133	12.7	1170	7	US-11-080-026-2	Sequence 2, Appli
44	133	12.7	1170	7	US-11-107-028-4	Sequence 4, Appli
45	128	12.2	1167	7	US-11-097-125-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-453-372-2
; Sequence 2, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789290
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining prior application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1

```
; SEQ ID NO 2
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-2

Alignment Scores:
Pred. No.: 1.63e-62 Length: 445
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 6 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-453-372-2 (1-445)
QY 1 GGGGACCTGATGTTCTCTGTGAGCAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCGG 60
   |||
Db 33 GlyspleuMePheLeuLeuAspSerSerAlaSerValSerHisTgIuPheSerArg 52
QY 61 GTTCGGAGATTGTGTGGGAGAGCTGTGAGCTCCACTGCCCCGAGCAGCCGGCCCTGCGT 120
   |||
Db 53 ValArgIuPheValGlyGluLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY 121 GCCACTGTGTGCACGTGAGGAGCTGCGCCATACACGAGTTCCCTTGGCCAGCAGCAGC 180
   |||
Db 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGAGATCGGGTGGTCTTCCGCGAGGATGGGTGACACCCAG 240
   |||
Db 93 SerGlyValAlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlyAspThrHis 112
QY 241 ACTGGCTGTGCGCTGTGATGCTCAAGAGAACAGCTGTTTGTGAAAGCATCAAGTCCCG 300
   |||
Db 113 ThrGlyLeuAlaLeuValTyrAlaGlyGluGlnLeuPheAlaGluAlaSerGlyAlaArg 132
QY 301 CCAAGGAGTCCCAAGAGTGTGTGTGAGTGCAGATGGCGGCTCCAGGAGCCCTGTGGC 360
   |||
Db 133 ProGlyValProGlyValLeuValTyrValThrAspGlyGlySerSerAspProValGly 152
QY 361 CCCCCCAGCAGAGCTCAAGAGCCTGGGCGCTCACCGGTTTCAATTGTGAGCAGCCGCGA 420
   |||
Db 153 ProPheMetGlnGluLeuLeuAspLeuGlyValThrValPheIleValSerThrGlyArg 172
QY 421 GGCAATTCTGTGAGCTGTCAAGCCGCTGCTCAGCCCTGCCAGAGACCTGTGACTTT 480
   |||
Db 173 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluIuHisLeuHisPhe 192
QY 481 GTGACGTGATGACTGTGACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
   |||
Db 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210

RESULT 2
US-11-186-284-26
; Sequence 26, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: THERAPY OF COLON CANCER
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
```

```
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3063
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-186-284-26

Alignment Scores:
Pred. No.: 2.2e-13 Length: 3063
Score: 261.50 Matches: 64
Percent Similarity: 52.5% Conservative: 29
Best Local Similarity: 36.2% Mismatches: 83
Query Match: 24.9% Indels: 1
DB: 7 Gaps: 1

US-10-699-035A-1 (1-537) x US-11-186-284-26 (1-3063)
QY 4 GACCTGATGTTCTGTGAGCAGCTCAGCCAGCTCTCTCACTACGACTTCTCCGGATT 63
   |||
Db 440 AspIleValPheLeuValAspGlySerTyrSerIleGlyIleAlaAsnPheValIleVal 459
QY 64 CGGAGATTGTGGGAGCAGCTGTGCTCCACTGCCCCCTGGGACACCGGCGCCCTGCTGCC 123
   |||
Db 460 ArgAlaPheLeuGluValLeuValLysSerPheGluIleSerProAsnArgValGlnIle 479
QY 124 AGTGTGTGCACGTGAGGAGCTGCGCATACACGAGTTCCCTTGGCCAGCAGACAGCTCG 183
   |||
Db 480 SerLeuValGlnTyrSerArgAspProHisThrGluPheThrLeuValYsPheThrLys 499
QY 184 GGTGAGCTGCCACAGATCGCGTGTGCTTCTGCCACCGCATGGGTGAACCCACACT 243
   |||
Db 500 ValGlnAspIleIleGluAlaIleAsnThrPheProTyrArgGlyGlySerThrAsnThr 519
QY 244 GGCCTGGCGCTGTGTTATGCCAAGAACAGCTGTTTGTGAAAGATCAGTGCCCGCGCA 303
   |||
Db 520 GlyLysAlaMetCnTyrValArgGluLysIlePheValProSerLysGlySerArgSer 539
QY 304 GGGGTGCCCAAGAGTGTGTGTGAGTGCAGATGGCGGCTCCAGACCCCTGTGGGCCCC 363
   |||
Db 540 AsnValProLysValMetIleLeuIleThrAspGlyLysSerSerAspAlaPheArgAsp 559
QY 364 CCATGACAGAGCTCAAGAGCCTGGGCGCTCACCGTTCATTGTCAAGACCGCCGAGGC 423
   |||
Db 560 ProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValLysAspAla 579
QY 424 AACTTCCTGAGCTGTACCGCGCTGCTCAGCCCTCGCGAGAGACCTGCACATTGTG 483
   |||
Db 580 ValArgSerGluLeuGluAlaIleAlaSerProAlaGluThrHisValPheThrVal 599
QY 484 --GACGTGATGACTGTGACATCATTTGTCCAAAGCTGAGGGGCTCCATT 531
   |||
Db 600 GluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIle 616

RESULT 3
US-11-169-041-160
; Sequence 160, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; CURRENT FILING DATE: 2005-06-28
```

```

; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 160
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-169-041-160

Alignment Scores:
Pred. No.: 4,94e-13 Length: 517
Score: 257.50 Matches: 64
Percent Similarity: 52.0% Conservative: 27
Best Local Similarity: 36.6% Mismatches: 83
Query Match: 24.5% Indels: 1
DB: Gaps: 1

US-10-699-035A-1 (1-537) x US-11-169-041-160 (1-517)
QY 10 ATGTTCTGCTGACAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCCGGTTCCGAG 69
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 MetPheLeuValAspIySerIySerIleGlyIleAlaAsnPhenValIyValAla 162
QY 70 TTGTGGGAGAGCTGTGCTCACTGCCCTGGGACCGGGGCGCTGCGCTG 129
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 PheLeuGluValIleuValIySerPheGluIleSerProAsnArgValGlnIleSerLeu 182
QY 130 GTGCACTGGGAGCTGGCCATACACGAGTTCCCTTGGCGCAGACAGCTCGGTG 189
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 ValGlnIySerArgAspProIstnGluPheThrLeuIySerPheThrIyValGlu 202
QY 190 GCTGCCAGAGTCGGTGGCTGCTTGGCCAGCGGAGGTGATACACCACTGGCTG 249
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 AspIleIleGluAlaIleAsnThrPheProIyArgGlyIySerThrAsnThrIyIyS 222
QY 250 GCGCTGCTTATGCCAAGAAGAGCTGTTCTGAAAGATCAGAGTCCCGGCGAGGGTG 309
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 AlaMetThrIyValArgIyIySlePheValProSerIySerIySerIySerIySer 242
QY 310 CCCAAGTCTGTGTGGGTGACAGATGGCGGCTCCAGGACCCCTGGGCCCCCGGAG 369
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 ProIyValMetIleuIleThrAspGlyIySerSerAspAlaPheArgAspProAla 262
QY 370 CAGAGCTCAAGAGCTGGGCGCTGACCGTTCATGTGTCAGACCGGCGAGCACTTC 429
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 IleIySleuArgAsnSerAspValGluIlePheAlaValGlyValIyAspAlaVal 282
QY 430 CTGAGCTGTCAAGCCGCTGCTGAGCCCTGCCGAGAGACCTGCACCTTGTG--GAC 486
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 SerGluLeuGluAlaIleAlaSerProProIaGluThrIleValPheThrValGlu 302
QY 487 GTGGATGACCTGCATCTTGTGCCAAGAGCTGAGGGGCTCCATT 531
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 PheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIle 317

RESULT 4
US-10-131-826A-294
; Sequence 294, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
```

```

; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 294
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-294

Alignment Scores:
Pred. No.: 3,34e-12 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: Gaps: 2

US-10-699-035A-1 (1-537) x US-10-131-826A-294 (1-915)
QY 4 GACCTGATGTTCTGCTGACAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCCGGTT 63
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 AspLeuValPheIleIleAspSerSerValAsnThrIleAspIyAlaIyVal 76
QY 64 CCGGAGTTGTGGGAGCTGTGCTCAGCTGCCCTTGGGACCGGGGCGCTGCC 123
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 IySleuPheIleValAspIleIleuGlnPheLeuAspIleGlyProAspValThrAspVal 96
QY 124 AGCTGTGTCAGCTGAGCTGCGGCATACACGAGTTCCCTTGGCGCAGACAGCTG 183
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 GlyLeuLeuGlnIyGlySerThrValIyAsnGluPheSerIleuIyThrPheIyAsp 116
QY 184 GGTGAGGCTGCCAGAGATGCGGTGCTTGTGCCAGCGAGTGGTACACCCACT 243
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 IySerGluValIglAlaValIyAspGluValIyAspGluValIySerThrIyThrMetThr 136
QY 244 GCGCTGGCGCTGTCTATGCCAAGAAGAGCTGTTGTCAGAGCTACAGTCCCGGCA 303
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 GlyLeuAlaIleGlnIyAlaLeuAsnIleAlaPheSerGluAlaGluIyAlaArgPro 156
QY 304 -----GGGATCCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 354
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 LeuArgGluAsnValProArgValIleMetIleValIlnAspGlyArgProGlnAspSer 176
QY 355 GTGGGCCCCCATGAGAGAGCTCAAGAGCTGGGGGCTCAGCTGTTCATTGTACAGACC 414
```

```
Db      177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
Qy      415 GCGCGAGGCACTTCCTGGAGCTGTCAGCCGCTGCTCAGCCGCGAGAGGACCTG 474
Db      197 GlyIleValAlaAspPheAsnThrLeuLysSerIleGlySerGluProHISGluAspHisVal 216
Qy      475 -----CACTTTGTGAGCGTGAGTACCTGCATCATTTGTCACAGAG 516
Db      217 PheIleuValAlaAsnPheserGlnIleGluThrLeuThrSerValPheGlnLys 234

RESULT 5
US-11-113-424-39
; Sequence 39, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gargolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 39
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-39

Alignment Scores:
Pred. No.:      3,34e-12      Length:      956
Score:          246.50      Matches:      57
Percent Similarity: 52.2%      Conservative: 36
Best Local Similarity: 32.0%      Mismatches: 78
Query Match:    23.5%      Indels:      7
DB:              7          Gaps:      2

US-10-699-035A-1 (1-537) x US-11-113-424-39 (1-956)
Qy      4 GACCTGATGTTCTCTGCTGAGACGCTCAGCCAGCTCTCTCACTACGAGTTCCCGGGTT 63
Db      57 AspLeuValAlaPheIleIleAspSerSerArgSerValAsnThrHisAspTyrAlaLysVal 76
Qy      64 CGGAGTTTGTGGGACAGCTGTGAGTCCACTGACCCCTGGAGCAGCGGGCCCTGCGTGC 123
Db      77 LysGluPheIleValAlaAspIleLeuGlnPheLeuAspIleGlyProAspValThrArgVal 96
Qy      124 AGTCGTGTGACGTGGGACGTGCGCCATACACCGAGTTCCCTTGGCGGACGACAGCTGC 183
Db      97 GlyLeuLeuGlnTyrGlySerThrValLysAsnGluPheSerLeuLysThrPheLysArg 116
Qy      184 GGTGAGGCTGCCAGAGATGCGGTGCTTCTGCCACAGGATGGTGACACCCACACT 243
Db      117 LysSerGluValGluArgAlaValLysArgMetArgHisLeuSerThrGlyThrMetThr 136
```

```
Qy      244 GCGCTGGCGCTGTCTATGCCAAGACAGCTTTTGTGAGCATCAGTGCCCGGCCA 303
Db      137 GlyLeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluAlaGluValAlaArgPro 156
Qy      304 -----GGGTCGCCCAAGTGTGCTGTGGGTGACAGTGGCGGTCCAGGACCT 354
Db      157 LeuArgGluAsnValProArgValIleValIleValAlaThrAspGlyArgProGlnAspSer 176
Qy      355 GTGGAGCCCCCAAGCAGACAGCTCAGAGCCTGGGCGGTCAACCGTTTCATTGTGACGACC 414
Db      177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
Qy      415 GCGCGAGGCACTTCCTGGAGCTGTCAAGCCGCTGCTCAGCCCTGCCGAGAGACCTG 474
Db      197 GlyIleValAlaAspPheAsnThrLeuLysSerIleGlySerGluProHISGluAspHisVal 216
Qy      475 -----CACTTTGTGAGCGTGAGTACCTGCATCATTTGTCACAGAG 516
Db      217 PheIleuValAlaAsnPheserGlnIleGluThrLeuThrSerValPheGlnLys 234

RESULT 6
US-11-192-449-6
; Sequence 6, Application US/11192449
; Publication No. US2005028181B1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Koreliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USC3
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-192-449-6

Alignment Scores:
Pred. No.:      2.3e-08      Length:      214
Score:          197.00      Matches:      65
Percent Similarity: 47.9%      Conservative: 28
Best Local Similarity: 33.5%      Mismatches: 81
Query Match:    18.8%      Indels:      20
DB:              7          Gaps:      7

US-10-699-035A-1 (1-537) x US-11-192-449-6 (1-214)
Qy      4 GACCTGATGTTCTCTGCTGAGACGCTCAGCCAGCTCTCTCACTACGAGTTCTCCGGTT 63
Db      22 AspIleValIleValAlaLeuAspGlySerAsnSerIle-----TyrProTyrAspSerVal 39
Qy      64 CGGAGTTTGTGGGACAGCTGTGAGTCCACTGACCCCTGGGACAGCGGGCCCTGCGTGC 123
Db      40 ThrAlaPheLeuAsnAspLeuLeuLysArgMetAspIleGlyProLysGlnThrGlnVal 59
Qy      124 AGTCGTGTGACGTGGGACGTGCGCCATACACCGAGTTCCCTTGGCGGACGACAGCTGC 183
Db      60 GlyIleValIleGlnTyrGlyGluAsnValThrHisGluPheAsnLeuAsnLysTyrSerSer 79
Qy      184 GGTGAGGCTGCCAGAGATGCGGTGCTTCTGCCACAGGATGGT--GACACCCAC 240
```

```

Db      80 ThrGluGluValLeuValAlaAlaLysLysIleValGlnArgGlyIleArgGlnThrMet 99
QY      241 ACTGGCGTGGCGCTGTATGCTATGCAAGAAACAGCTTTGTCGAACATCAGGTGCCGG 300
Db      100 ThrAlaLeuGlyThrAspThrAlaArgLysGluAlaPheThrGluAlaArgGlyAlaArg 119
QY      301 CCAGGGGTGCCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 354
Db      120 ArgGlyValAlaLysLysValMetValIleValIleThrAspGlyIleSerHisAspAsnHisArg 139
QY      355 GTGGGCCCCCATGAGAGAGCTCAAGAGCTGGGGGTACCGCTTTCATTGTTC----- 408
Db      140 LeuLysLysValIleGlnAspCysGluAspGluAsnIleGlnArgPheSerIleAlaIle 159
QY      409 -----AGACCGCGCCGAGCAACTTCCTG-----GAGCTGTACGCC 444
Db      160 LeuGlySerTyrAsnArgLysAsnLeuSerThrGluLysPheValGluGluIleLysSer 179
QY      445 GCTGCTCAGCCCTGCGCGAGAACGACCTGACTTTGTGAGCTGTGATGAC----- 495
Db      180 IleAsnSerGluProThrGluLys-----HisPheAsnValSerAspGluLeuAla 197
QY      496 CTGCACATCATTTGTCCAAAGAGCTGAGGGGCTCCATTCTGCG 537
Db      198 LeuValThrIleValLysThrLeuGlyIleArgIlePheAla 211

RESULT 7
US-11-192-449-9
; Sequence 9, Application US/11192449
; Publication No. US2005028181A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCN3
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-192-449-9

Alignment Scores:
Pred. No.: 2,3e-08 Length: 214
Score: 197.00 Matches: 65
Percent Similarity: 47.9% Conservative: 28
Best Local Similarity: 33.5% Mismatches: 81
Query Match: 18.8% Indels: 20
DB: 7 Gaps: 7

US-10-699-035A-1 (1-537) x US-11-192-449-9 (1-214)
QY      4 GACCGATGTCCTGCTGTGACAGCTCAGCGAGGTCTCTCACTAGAGATTCTCCGGGTT 63
Db      22 AapIleValIleValLeuAspGlySerAsnSerIle-----TyrProTrpAspSerVal 39
QY      64 CGGAGTTTGTGGGCGAGCTGTGCTCCACTGCCCGGAGCCCGGAGCCCTGCGTGC 123

```

```

Db      40 ThrAlaPheLeuAsnAspLeuLysArgMetAspIleGlyProLysGlnThrGluVal 59
QY      124 ACTGTGTCACAGTGGGCGAGTCCGACATACACGAGTTCCCTCGCCAGACAGCTCG 183
Db      60 GlyIleValGlnTyrGlyGluAsnValThrHisGluPheAsnLeuAsnLysTyrSerSer 79
QY      184 GGTAGGCTGCGCCAGAGTCCGTCGCTTGTGCCCGACGAGTGGT--GACACCCAC 240
Db      80 ThrGluGluValLeuValAlaAlaLysLysIleValGlnArgGlyIleArgGlnThrMet 99
QY      241 ACTGGCGTGGCGCTGTATGCTATGCAAGAAACAGCTTTGTCGAACATCAGGTGCCGG 300
Db      100 ThrAlaLeuGlyThrAspThrAlaArgLysGluAlaPheThrGluAlaArgGlyAlaArg 119
QY      301 CCAGGGGTGCCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 354
Db      120 ArgGlyValAlaLysLysValMetValIleValIleThrAspGlyIleSerHisAspAsnHisArg 139
QY      355 GTGGGCCCCCATGAGAGAGCTCAAGAGCTGGGGGTACCGCTTTCATTGTTC----- 408
Db      140 LeuLysLysValIleGlnAspCysGluAspGluAsnIleGlnArgPheSerIleAlaIle 159
QY      409 -----AGACCGCGCCGAGCAACTTCCTG-----GAGCTGTACGCC 444
Db      160 LeuGlySerTyrAsnArgLysAsnLeuSerThrGluLysPheValGluGluIleLysSer 179
QY      445 GCTGCTCAGCCCTGCGCGAGAACGACCTGACTTTGTGAGCTGTGATGAC----- 495
Db      180 IleAsnSerGluProThrGluLys-----HisPheAsnValSerAspGluLeuAla 197
QY      496 CTGCACATCATTTGTCCAAAGAGCTGAGGGGCTCCATTCTGCG 537
Db      198 LeuValThrIleValLysThrLeuGlyIleArgIlePheAla 211

RESULT 8
US-11-192-449-5
; Sequence 5, Application US/11192449
; Publication No. US2005028181A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCN3
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Rat
US-11-192-449-5

Alignment Scores:
Pred. No.: 1,13e-07 Length: 214
Score: 188.00 Matches: 62
Percent Similarity: 48.5% Conservative: 32
Best Local Similarity: 32.0% Mismatches: 80
Query Match: 17.9% Indels: 20
DB: 7 Gaps: 7

US-10-699-035A-1 (1-537) x US-11-192-449-5 (1-214)

```

```

QY      4 GACCTGATGTTCTCTGTCGACAGCTCAGCAGCTCTCTCACTACGAGTTCTCCCGGTT 63
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      22 Aspllevalillevalilleaspolyseranserle-----tyrprotrpcluserval 39
QY      64 CGGAGTTTGTGGGACAGCTGTGTGCTCCACTGCCCTGGACCCGGGCCCCCTGGTCC 123
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      40 Ilealpheleuanasplleuileuylsargmetasplilegilyprolysglntthcglinal 59
QY      124 AGCTGTGGCACTGTGGCACTGTGGCACTGTGGCACTGTGGCACTGTGGCACTGTGGCA 183
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      60 Glyllevalglntryglylualsnvalthrhisglupheanleuanmlystyserser 79
QY      184 GGTAGGCTGCCAGATGCGGTCGTGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 240
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      80 Thriglualleuvalleuallealalyvalilegilyarglinsgilyleuglntthmet 99
QY      241 ACTGAGCTGTGGCTGTGTATGCAAGACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      100 Thrilaleuglylileaspthrilaaargylsualaphethrglualaaargglyalarg 119
QY      301 CCAGGGGGCCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 354
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      120 Argglyvallylsvalmetvalillevalthrapsglyluserhiaspsmtyrarg 139
QY      355 GTGGGCCCCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 414
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      140 Leuylsglnvalilleglnaspcysgluaspeluanenileglnargphserillealale 159
QY      415 -----GGCCGAGGCACTTCTGTG-----GAGCTGTACGCC 444
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      160 Leugllyhistryaasnarglyasneuserthrghlulysphevalgluglntleyseser 179
QY      445 GCTGCTGAGGCCCTGCGGAGAGCACTGCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 495
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      180 Ilelaasergluprothghlulys-----hisphenearvalseraspgluleuala 197
QY      496 CTGCACATCATTTGTCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 517
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      198 LeuvalThrillevallyalaleuclglunrgrillepheala 211

RESULT 9
US-10-063-703-34
; Sequence 34, Application US/10063703
; Publication No. US20060008901a1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,703
; PRIORITY FILING DATE: 2002-05-08
; PRIORITY APPLICATION REMOVED - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 34
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-703-34

Alignment Scores:
Pred. No.: 2,55e-07 Length: 678
Score: 183.00 Matches: 53
Percent Similarity: 45.1% Conservative: 25
Best Local Similarity: 30.6% Mismatches: 91

```

```

Query Match: 17.4% Indels: 4
DB: 6 Gaps: 2
US-10-699-035A-1 (1-537) x US-10-063-703-34 (1-678)

QY      4 GACCTGATGTTCTCTGTCGACAGCTCAGCAGCTCTCTCACTACGAGTTCTCCCGGTT 63
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      495 Aspllegllyphevalilleasptgylseranservaldlythrghlyanpheargthrval 514
QY      64 CGGAGTTTGTGGGACAGCTGTGTGCTCCACTGCCCTGGACCCGGGCCCCCTGGTCC 123
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      515 Leuglnphevalthransleuthlysgluphegluliseaspthrmrappthrargille 534
QY      124 AGCTGTGGCACTGTGGCACTGTGGCACTGTGGCACTGTGGCACTGTGGCACTGTGGCACT 183
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      535 Glylavalglntrythrtryglylualargluegluphegllypheasplystyserser 554
QY      184 GGTAGGCTGCCAGATGCGGTCGTGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 243
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      555 Lysproasplleleuanalallelyasargvalglytyrtrpserdlyglythrsertthr 574
QY      244 GGCTGGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 303
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      575 Glylalaalleasnphealaleuclglnleuhenlyls-----serlyspbro 591
QY      304 GGGGTGCCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 363
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      592 Asnlyasrlylsleumetilleuleillethrapslyrgerstlyraaspsvalargille 611
QY      364 CCATGTGAGAGCTCAAGAGCACTGTGGCTCACCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 423
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      612 Proalamealalaahisleuylsaglyvalillethryralalleglyvalailetrpala 631
QY      424 AACTCTGTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 483
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      632 Alaeglntgluleglualillealathrisproalaaagaprhissertphepheval 651
QY      484 GAC---GTGATGACCTGTGCATCATTTGTCCAGAGCTG 519
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      652 Asplupheaspsenleuhsieglntryvalproargille 664

RESULT 10
US-11-102-240-34
; Sequence 34, Application US/11102240
; Publication No. US20050260647A1
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: ANTIODIES TO POLYPEPTIDES ENCODED BY A NUCLEIC ACID UNDEREXPRES
; FILE REFERENCE: P3230R1C106C
; CURRENT APPLICATION NUMBER: US/11/102,240
; PRIORITY FILING DATE: 2005-04-08
; PRIORITY APPLICATION NUMBER: 10/063662
; PRIORITY FILING DATE: 2002-05-07
; PRIORITY APPLICATION NUMBER: 10/006867
; PRIORITY FILING DATE: 2001-12-06
; PRIORITY APPLICATION NUMBER: PCT/US00/23328
; PRIORITY FILING DATE: 2000-08-24
; PRIORITY APPLICATION NUMBER: 60/170262
; PRIORITY FILING DATE: 199-12-09
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 34
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-102-240-34

Alignment Scores:
Pred. No.: 2,55e-07 Length: 678

```


Score:	183.00	Matches:	53
Percent Similarity:	45.1%	Conservative:	25
Best Local Similarity:	30.6%	Mismatches:	91
Query Match:	17.4%	Indels:	4
DB:	7	Gaps:	2

US-10-699-035A-1 (1-537) X US-11-102-240-34 (1-678)

QY	GACCTATGTTCTTCCTGGACAGCTCAGCCAGCTCTCTCACTAAGATTTCTCCGGGTT	63
Db	495 AsplleclgYlPhevalIleapglYserSerSerValGlythrlgYasmpheArgThnVal	514
QY	64 CGGAGATTGTGGGCGACGTGGTGCCTCCACTGCCCTCGGGACCGGGGCTGCGTCC	123
Db	515 LengllPhevalIThrasnleuThrlysglupheglulleSerapThraspThraArgIle	534
QY	124 AGTCTGGTGCACGTGGGCGAGTGGCCATACACGAGTTCCCTTGCGGCACACAGCTCG	183
Db	535 GYlYAlaValGlnTyrThrYgluGlnAlaYgluGlnupheglYpheaSpYsYsSerSer	554
QY	184 GGTGAGCTGCCCGAGATGGGTGGCTTCTGCCACAGCGATGGTGACACCCACT	243
Db	555 LysProaspIleuasnAlaIleuYsaYgValGlyTyrTrpserGlyThrSerThr	574
QY	244 GGCCTGGCGGTGTCTATGCAAGAAACAGCTGTTCGTAAGCATACAGTGGCCGGGCA	303
Db	575 GYlYAlaIleIleasnPheAlaLeuGlnGlnleuPheYsYs-----SerLysPro	591
QY	304 GGGGTGCCCAAGTGTGTGTGGTGGGTGACAGATGGCGGCTCCAGGACCCGTGGGCCC	363
Db	592 AsnlyeArgYsleuMetIleleuIleThrAspGlyArgSerTyrAspAspValArgIle	611
QY	364 CCCATGACAGAGCTCAAGACCTGGGGCTCACCGGTTCATTGTTCAGACCGGGCGAGGC	423
Db	612 ProIaMetAlaIaIaHisleuYsgYsValIleThrYrAlaIleGlyValaIaIaTrpAla	631
QY	424 AACCTCTGAGCTGTTCAGCCGCTGGCTCAGCCCTCGCGGAAGACACTTCATTGTTG	483
Db	632 AlAGlnGlnGlnleuGlnValIleAlaIaThrHisproIaIaArgAspHisSerPhePheVal	651
QY	484 GAC---GTGATGACCTGCACATCATTTGTCCAAGAGCTG	519
Db	652 AspGlnPheaSpasnleuHisGlnTyrValProArgIle	664

RESULT 11

US-11-080-026-4
; Sequence 4, Application US/11080026
; Publication No. US20050260192A1
; Journal Information

```

1  GENERAL INFORMATION:
2  APPLICANT: Springer, Timothy A.
3  APPLICANT: Shimooka, Motomu
4  APPLICANT: Lu, Chafen
5  TITLE OF INVENTION: MODIFIED POLYPEPTIDES STABILIZED IN A
6  TITLE OF INVENTION: DESIRED CONFORMATION AND METHODS FOR PRODUCING SAME
7  FILE REFERENCE: CFBF-P02-021
8  CURRENT APPLICATION NUMBER: US/11/080,026
9  CURRENT FILING DATE: 2005-03-15
10 PRIOR APPLICATION NUMBER: 09/945,265
11 PRIOR FILING DATE: 2001-08-31
12 PRIOR APPLICATION NUMBER: US 60/229,700
13 PRIOR FILING DATE: 2000-09-01
14 NUMBER OF SEQ. ID NOS: 4
15 SOFTWARE: FastSeq for Windows Version 4.0
16 SEQ. ID NO. 4
17 LENGTH: 1152
18 TYPE: PRT
19 ORGANISM: Homo sapiens
20 US-11-080-026-4

```

[illegible]

Best Local Similarity:	28.3%	Mismatches:	92
Query Match:	15.9%	Indels:	11
Ds:	7	Gaps:	5

US-10-699-035A-1 (1-537) x US-11-080-026-4 (1-1152)

US-10-699-035A-1 (1-537) x US-11-080-026-4 (1-1152)

Oy	GACCTATGTTCTGCTGGACAGCTCAGCAGCGCTCTCATGCAAGTCTCCGGGTT	63
Db	150 AsplleAlaPheLeuLeuIlePolyserylSerIleIleProIleAspPheAlaArgMet	163
Oy	64 CGGAGATTGTGGGAGCAGCTGTGACTCCACTGGCCCGGACCGGGCCCTGCTGCC	122
Db	170 LysGluPheValSerThrValMetGluGlnLeuLysLysSerLysThrLeuPhe-----	187
Oy	124 AGCTGTGTCACGTGGGACAGTGGCCATACACCGAGTTCCCTTGTGGCCAGACAGCTCG	183
Db	188 SerLeuMetGlnTyrSerGluGluPheArgIleHisPheThrPheLysGluPheGlnAsn	207
Oy	184 GGTGAGGCTGCCAGGATCGGTGGGTGCTTCTGCCAGCGATGGGGACACCCACT	245
Db	208 AsnProAsnProArgSerLeuValLysProIleThrGlnLeuLeuGlyValArgThrHisThr	222
Oy	244 GGCCTGCGCTGTCTATGCCAAGGACAGCTGTTTGCTGAAGCATCAGTGCCTGGCCA	303
Db	228 AlaThrGlyIleArgLysValValArgGluLeuPheAsnIleThrAsnGlyAlaArgLys	247
Oy	304 GGGGTGCCCAAGTCTGTGTGGGTGACAGATGGCGGCTCC--AGCGACCTGTGGGC	366
Db	248 AsnAlaPheLysIleLeuValValIleThrAspGlyGluLysPheGlyAspProLeuGly	267
Oy	361 -----CCCCCATGAGAGCTCAAGACCTGGGGCGTCACCGTTCATTGTAGGACACC	411
Db	268 TyrGluAspValIleProGluIleAspArgGluGlyValIleArgTyrValIleGlyVal	287
Oy	415 GGC-----CGAGCAACTTCTCGAGCTGTGACCGCTGCTCAGCCCT	455
Db	288 GlyAspAlaPheArgSerGluLysSerArgGlnLeuAsnThrIleAlaSerLysPro	307
Oy	460 GCCGAGAGACCTGCATTTGTG--GACGTGATACCTGCACATCATTTGTCCAAAG	511
Db	308 ProArgAspHisValPheGlnValAlaAsnAspMetGluAlaLeuLysTyrIleGlnAsnGln	327
Oy	517 CTGAGGGGCTCCATTCTCGCG	537
Db	328 LeuArgGluLysIlePheAla	334

RESULT 12

US-10-453-372-186
; Sequence 186, Application US/10453372
; Publication No. US20060003323A1
GENERAL INFORMATION:

```

1  /  APPLICANT: Alsbjorn, et al.
2  /  TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
3  /  FILE REFERENCE: 21402-589 A
4  /  CURRENT APPLICATION NUMBER: US/10/453,372
5  /  PRIOR FILING DATE: 2003-06-03
6  /  PRIOR APPLICATION NUMBER: 09/7893390
7  /  PRIOR FILING DATE: 2001-02-23
8  /  PRIOR APPLICATION NUMBER: 60/185967
9  /  PRIOR FILING DATE: 2000-03-01
10 /  PRIOR APPLICATION NUMBER: 09/823187
11 /  PRIOR FILING DATE: 2001-03-29
12 /  PRIOR APPLICATION NUMBER: 60/195792
13 /  PRIOR FILING DATE: 2000-03-10
14 /  PRIOR APPLICATION NUMBER: 09/839446
15 /  PRIOR FILING DATE: 2001-03-19
16 /  PRIOR APPLICATION NUMBER: 60/199476
17 /  PRIOR FILING DATE: 2000-03-25
18 /  PRIOR APPLICATION NUMBER: 09/863776
19 /  PRIOR FILING DATE: 2001-05-23
20 /  PRIOR APPLICATION NUMBER: 60/208263
21 /  PRIOR FILING DATE: 2000-05-31
22 /  PRIOR APPLICATION NUMBER: 09/939398

```

```
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: 60/227800
/ PRIOR FILING DATE: 2000-08-25
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ SOFTWARE: Curation version 0.1
/ SEQ ID NO: 186
/ LENGTH: 709
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-453-372-186

Alignment Scores:
Pred. No.: 1.04e-05 Length: 709
Score: 162.00 Matches: 53
Percent Similarity: 47.1% Conservative: 27
Best Local Similarity: 31.2% Mismatches: 74
Query Match: 15.4% Indels: 16
DB: 6 Gaps: 5

US-10-699-035A-1 (1-537) x US-10-453-372-186 (1-709)

OY 4 GACCGATGTTCTCTGAGACGCTCAGCCAGGCTCTCTCACTAGAGTTCTCCGGGTT 63
DB 68 GluLeuValPheLeuValAspSerSerSerValGlyGluValAsnPhaArgSerGlu 87
OY 64 CGGAGATTGTGGGAGAGCTGTGCTCCACTGCGCCCTGGGAGCCGGGCGCTGCGTCC 123
DB 88 LeuMetPheValArgGlyLeuLeuSerAspPheProValValProThrAlaThrArgVal 107
OY 124 AGTCTGTGCACGTGGGAGCTGCGCATACACCGAGTTCCCTTGGCCGACGACAGCTCG 183
DB 108 AlaIleValThrPheSerSerIleAsnTyValValProArgAlaAspTyrIleSer 126
OY 184 GGTAGGCTGCCAG-----GATCGGCTGCTCTTCTGCC 219
DB 127 ThrArgAlaArgGlnHisIleSylalaLeuLeuGlnGluIleProAlaIleSer 146
OY 220 CAGCGATGTGGAGACCCACACTGCGCTGCTGTATGCCAAGAGACGCTGTT 279
DB 147 TyrArgGlyGlyGlyThrIleSylalaPheGlnGlnAlaGlnIleLeu 166
OY 280 GCTGAGCATCAGGTGCGGCGGCTGCCAAAGTCTGTGCTGCTGAGATGCG 339
DB 167 -----HisAlaArgGluAsnSerAlaIleValValPheLeuIleThrAspGly 182
OY 340 GGTCTCAG-----GACCTGTGGGCCCCCATGACGAGACTCAAGGACTGCGCTC 393
DB 183 TyrSerAsnGlyGlyAspPro--ArgProIleAlaIleAsnSerLeuArgAspSerGlyVal 201
OY 394 ACCGATGTCATGTGAGACCGGAGGCACTTCTGTGAGCTGTGAGCCGCTGCTCA 453
DB 202 GluIlePheThrPheGlyIleTrrGlnGlyAsnIleArgGluLeuAsnMetAlaSer 221
OY 454 GCCCTGCCGAGAGCACTGCACTTTGTG 483
DB 222 ThrProGlyGluHisCysTyrLeuLeu 231

RESULT 13
US-10-453-372-180
/ Sequence 180, Application US/10453372
/ Publication No. US20060003323A1
/ GENERAL INFORMATION:
/ APPLICANT: Alcobrook, et al.
/ TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
/ FILE REFERENCE: 21402-589 A
/ CURRENT APPLICATION NUMBER: US/10/453,372
/ PRIOR FILING DATE: 2003-06-03
/ PRIOR APPLICATION NUMBER: 09/789390
/ PRIOR FILING DATE: 2001-02-23
/ PRIOR APPLICATION NUMBER: 60/185967
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: 09/823187
```

```
/ PRIOR FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/195792
/ PRIOR FILING DATE: 2000-03-10
/ PRIOR APPLICATION NUMBER: 09/839446
/ PRIOR FILING DATE: 2001-03-19
/ PRIOR APPLICATION NUMBER: 60/199476
/ PRIOR FILING DATE: 2000-03-25
/ PRIOR APPLICATION NUMBER: 09/863776
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: 60/208263
/ PRIOR FILING DATE: 2000-05-31
/ PRIOR APPLICATION NUMBER: 09/939398
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: 60/227800
/ PRIOR FILING DATE: 2000-08-25
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1609
/ SOFTWARE: Curation version 0.1
/ SEQ ID NO: 180
/ LENGTH: 709
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-453-372-180

Alignment Scores:
Pred. No.: 1.25e-05 Length: 709
Score: 161.00 Matches: 53
Percent Similarity: 46.7% Conservative: 26
Best Local Similarity: 31.4% Mismatches: 76
Query Match: 15.3% Indels: 14
DB: 6 Gaps: 4

US-10-699-035A-1 (1-537) x US-10-453-372-180 (1-709)

OY 4 GACCGATGTTCTCTGAGACGCTCAGCCAGGCTCTCTCACTAGAGTTCTCCGGGTT 63
DB 68 GluLeuValPheLeuValAspSerSerSerValGlyGluValAsnPhaArgSerGlu 87
OY 64 CGGAGATTGTGGGAGAGCTGTGCTCCACTGCGCCCTGGGAGCCGGGCGCTGCGTCC 123
DB 88 LeuMetPheValArgGlyLeuLeuSerAspPheProValValProThrAlaThrArgVal 107
OY 124 AGTCTGTGCACGTGGGAGCTGCGCATACACCGAGTTCCCTTGGCCGACGACAGCTCG 183
DB 108 AlaIleValThrPheSerSerIleAsnTyValValProArgValAspTyrIleSerThr 127
OY 184 GGTAGGCTGCCAG-----GATCGGCTGCTCTTCTGCCAG 222
DB 128 ArgArgAlaArgGlnHisIleSylalaLeuLeuGlnGluIleProAlaIleSerThr 147
OY 223 CGCATGAGTGAACCCACACTGCGCTGCTGTATGCCAAGAGACGCTGTTGCT 282
DB 148 ArgGlyGlyGlyThrIleSylalaPheGlnGlnAlaGlnIleLeu 166
OY 283 GAGCATGAGTGCCTGCGGCGGCTGCCAAAGTCTGTGCTGCTGAGTCAAGTGGCGGC 342
DB 167 -----HisAlaArgGluAsnSerThrIleValValPheLeuIleThrAspGlyTyr 183
OY 343 TCCAGC-----GACCTGTGGGCCCCCATGACGAGCTCAAGGACTGCGGCTCAC 396
DB 184 SerAsnGlyGlyAspPro--ArgProIleAlaIleAsnSerLeuArgAspSerGlyValGlu 202
OY 397 GTGTCATGTGTGAGACCGGAGGCACTTCTGTGAGCTGTGAGCCGCTGCTCAGCC 456
DB 203 IlePheThrPheGlyIleTrrGlnGlyAsnIleArgGluLeuAsnMetAlaSerThr 222
OY 457 CTCGCCGAGAGCACTGCACTTTGTG 483
DB 223 ProGlyGluHisCysTyrLeuLeu 231

RESULT 14
US-10-453-372-194
/ Sequence 194, Application US/10453372
```

```
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuiSeqlist version 0.1
; SEQ ID NO 194
; LENGTH: 3568
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-194

Alignment Scores:
Pred. No.: 1.14e-05 Length: 3568
Score: 161.00 Matches: 53
Percent Similarity: 46.7% Conservative: 26
Best Local Similarity: 31.4% Mismatches: 76
Query Match: 15.3% Indels: 14
DB: Gaps: 4

US-10-699-035A-1 (1-537) x US-10-453-372-194 (1-3568)

QY 4 GACCTGATGTTCTCTGCTGACAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCGGGTT 63
Db ::::::::::::::::::::::::::::
80 GlutLeuValPheLeuValAspAspSerSerValGlyGluValAsnPhaArgSerGlu 99
QY 64 CGGAGTTTGGGGCAGCTGGTGGCTCAGTGCCTGGGACCGGGGCGCTGGTGCC 123
Db ::::::::::::::::::::::::::::
100 LeuMetPheValArgGlyLeuLeuSerAspPheProValValProThrAlaThrArgVal 119
QY 124 AGCTGTGTCAGCTGGGACGTCGCCATACACCGAGTTCCCTTGGCCAGACAGCTCG 183
Db ::::::::::::::::::::::::::::
120 AlaIleValThrPheSerSerLysAsnTyValValProArgValAspTyrIleSerThr 139
QY 184 GGTGAGGCTGCCAG-----GATGCGGTGCGTCTTCTGCCAG 222
Db ::::::::::::::::::::::::::::
140 ArgArgAlaArgGlnHisLysCysAlaLeuLeuGlnGlnIleProAlaIleSerTyr 159
QY 223 CGCATGGGTGACACCACTGAGCTGGCGCTGTCTATGCAAGAAACAGCTGTTGTT 282
Db ::::::::::::::::::::::::::::
160 ArgGlyGlyValGlyThrTyrThrLysGlyAlaPheGlnGlnAlaGlnIleLeuLeu--- 178
QY 283 GAAGCATCAGGTGCGCCGCGCAGGGGTGCCAAAGTGTGGTGGGTGACAGATGGCGGC 342
Db ::::::::::::::::::::::::::::
179 -----HisAlaArgGlnLysSerThrLysValValPheLeuIleThrAspGlyTyr 195
QY 343 TTCAGC-----GACCTGTGGGCGCCCGCCATGACGAGACTCAAGGACCTGGCGCTCAC 396
Db ::::::::::::::::::::::::::::
196 SerAsnGlyGlyAspPro---ArgProIleAlaAlaSerLeuArgAspSerGlyValGln 214
```

```
QY 397 GTGTTCAATGTCAGCACCGCGGCAACTTCTGAGCTGTCTACGCCGCTGCACCC 456
Db ::::::::::::::::::::::::::::
215 IlePheThrPheGlyIleTyrPdnGlyAsnIleArgGlnLeuAsnAspMetAlaSerThr 234
QY 457 CTGCGCGAAGACACCTTCACCTTGTG 483
Db ::::::::::::::::::::::::::::
235 ProLysGlnGlnHisCysTyrLeuLeu 243

RESULT 15
US-10-453-372-178
; Sequence 178, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuiSeqlist version 0.1
; SEQ ID NO 178
; LENGTH: 3570
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-178

Alignment Scores:
Pred. No.: 1.14e-05 Length: 3570
Score: 161.00 Matches: 53
Percent Similarity: 46.7% Conservative: 26
Best Local Similarity: 31.4% Mismatches: 76
Query Match: 15.3% Indels: 14
DB: Gaps: 4

US-10-699-035A-1 (1-537) x US-10-453-372-178 (1-3570)

QY 4 GACCTGATGTTCTCTGCTGACAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCGGGTT 63
Db ::::::::::::::::::::::::::::::::::::
80 GlutLeuValPheLeuValAspAspSerSerValGlyGluValAsnPhaArgSerGlu 99
QY 64 CGGAGTTTGGGGCAGCTGGTGGCTCAGTGCCTGGGACCGGGGCGCTGGTGCC 123
Db ::::::::::::::::::::::::::::::::::::
100 LeuMetPheValArgGlyLeuLeuSerAspPheProValValProThrAlaThrArgVal 119
QY 124 AGCTGTGTCAGCTGGGACGTCGCCATACACCGAGTTCCCTTGGCCAGACAGCTCG 183
Db ::::::::::::::::::::::::::::::::::::
120 AlaIleValThrPheSerSerLysAsnTyValValProArgValAspTyrIleSerThr 139
QY 184 GGTGAGGCTGCCAG-----GATGCGGTGCGTCTTCTGCCAG 222
Db ::::::::::::::::::::::::::::::::::::
140 ArgArgAlaArgGlnHisLysCysAlaLeuLeuGlnGlnIleProAlaIleSerTyr 159
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 13:10:29 ; Search time 9.08696 Seconds
(without alignments)
603.637 Million cell updates/sec

Title: US-10-699-035A-6

Perfect score: 2154
Sequence: 1 MLPMTALGLALSLRLALARS.....RRPRPPVAPAPPGTASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 1312538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New:*
1: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubppaa/US10_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubppaa/US11_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2130.5	98.9	445	6	US-10-453-372-2
2	421.5	19.6	3063	7	US-11-186-284-26
3	401	18.6	517	7	US-11-169-041-160
4	253.5	11.8	915	6	US-10-131-826A-294
5	253.5	11.8	956	7	US-11-113-424-39
6	194	9.0	214	7	US-11-192-449-6
7	194	9.0	214	7	US-11-192-449-9
8	185	8.5	214	7	US-11-192-449-5
9	184	8.5	678	6	US-10-063-703-34
10	184	8.5	678	6	US-11-102-240-34
11	175	8.1	709	6	US-10-453-372-186
12	174	8.1	709	6	US-10-453-372-180
13	174	8.1	1152	7	US-11-080-026-4
14	173	8.0	709	6	US-10-453-372-182
15	172.5	8.0	1167	6	US-10-601-368-18
16	169	7.8	3570	6	US-10-453-372-178
17	169	7.8	3570	6	US-10-453-372-196
18	169	7.8	3570	6	US-10-453-372-198
19	169	7.8	3570	6	US-10-453-372-200
20	169	7.8	3570	6	US-10-453-372-202
21	169	7.8	3570	6	US-10-453-372-204
22	169	7.8	3570	6	US-10-453-372-206
23	168	7.8	709	6	US-10-453-372-184
24	168	7.8	3568	6	US-10-453-372-194
25	162.5	7.5	847	6	US-10-995-561-634

26	162.5	7.5	1259	6	US-10-995-561-625	Sequence 625, App
27	162.5	7.5	1286	6	US-10-995-561-628	Sequence 628, App
28	162.5	7.5	1341	6	US-10-995-561-621	Sequence 621, App
29	162.5	7.5	2217	7	US-11-193-561-38	Sequence 38, App1
30	162.5	7.5	2217	7	US-11-193-561-38	Sequence 38, App1
31	162.5	7.5	2217	7	US-11-193-789-38	Sequence 38, App1
32	162.5	7.5	2217	7	US-11-193-806-38	Sequence 38, App1
33	162.5	7.5	2217	7	US-11-193-857-38	Sequence 38, App1
34	162.5	7.5	2330	7	US-11-193-561-21	Sequence 21, App1
35	162.5	7.5	2330	7	US-11-193-771-21	Sequence 21, App1
36	162.5	7.5	2330	7	US-11-193-789-21	Sequence 21, App1
37	162.5	7.5	2330	7	US-11-193-806-21	Sequence 21, App1
38	162.5	7.5	2330	7	US-11-193-857-21	Sequence 21, App1
39	162.5	7.5	2355	6	US-10-995-561-623	Sequence 623, App
40	162.5	7.5	2355	6	US-10-995-561-627	Sequence 627, App
41	162.5	7.5	2355	7	US-11-193-561-19	Sequence 19, App1
42	162.5	7.5	2355	7	US-11-193-771-19	Sequence 19, App1
43	162.5	7.5	2355	7	US-11-193-789-19	Sequence 19, App1
44	162.5	7.5	2355	7	US-11-193-806-19	Sequence 19, App1
45	162.5	7.5	2355	7	US-11-193-857-19	Sequence 19, App1

ALIGNMENTS

RESULT 1
US-10-453-372-2
; Sequence 2, Application US/10453372
; Publication No. US200060003323A1
GENERAL INFORMATION:
; APPLICANT: Alsbrock, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 2
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-2

Query Match 98.9%; Score 2130.5; DB 6; Length 445;
Best Local Similarity 93.9%; Pred. No. 3.9e-161;
Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

QY 1 MLPMTALGLALSLRLALARS...RRPRPPVAPAPPGTASREP 60
DB 1 MLPMTALGLALSLRLALARS...RRPRPPVAPAPPGTASREP 60
QY 61 VAPLPGLTALRASLVHVGSRPYTEFPFGQSSGGEAODAVRASQRMGDTHGLALVYA 120

```

Db      61 VAPLPLGTALRASLVHVGSRPTTPEFQOHSSGEAQAQVRAQRMGDTHGLAVYA 120
Qy      121 KEOLFAEASGARGPVKVLVWTDGSSDPVGPMPQELKDLGVTVIVTSGRGNFLELSA 180
Db      121 KQOLFABAGARGPVKVLVWTDGSSDPVGPMPQELKDLGVTVIVTSGRGNFLELSA 180
Qy      181 MASAPAEKHLHVDVDDHIIYOELRGSILDMARPOQLHATEITSSGFLAMPPLITADS 240
Db      181 MASAPAEKHLHVDVDDHIIYOELRGSILDMARPOQLHATEITSSGFLAMPPLITADS 240
Qy      241 GYVVELVPSAQAARQQLPQNATDWIAGLDPTDYLVALVPSNRLRLPOLLRAR 300
Db      241 GYVVELVPSAQAARQQLPQNATDWIAGLDPTDYLVALVPSNRLRLPOLLRAR 300
Qy      301 TR-----PEAGPERIVISHARPSRLSVNAPALGSAA 333
Db      301 TRGEGAGPGASGPEGSGAPPTQALALPAPPEAGPERIVISHARPSRLSVNAPALGSAA 360
Qy      334 ALGYHVOGFPRLRGGEAQRVEVPAGRNCTTLOGIAPGTAVLVVTAAFRSGRESALSAAK 393
Db      361 ALGYHVOGFPRLRGGEAQRVEVPAGRNCTTLOGIAPGTAVLVVTAAFRSGRESALSAAK 420
Qy      394 TPDGPRPRRPVRAAPTGTASREP 418
Db      421 TPDGPRPRRPVRAAPTGTASREP 445

```

```

RESULT 2
US-11-186-284-26
; Sequence 26, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Millenium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Schlegel, Robert
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPW01-029P2RM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3063
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-26

```

```

Query Match      19.6%; Score 421.5; DB 7; Length 3063;
Best Local Similarity 27.5%; Pred. No. 5,5e-25;
Matches 128; Conservative 59; Mismatches 174; Indels 105; Gaps 9;
Qy      32 RQDLMLDLSASVSHYEFSSRVREYQGLVAPLPLGTALRASLVHVGSRPYTEPFQGH 91
Db      438 KADIVFLVDGSGYIGIANFVKRAFLFVVKSFELSPNRRVOISLVOYSRDPTHEFLTKAF 497
Qy      92 SSGEAAQADVRAQAQMGDTHGLALVYAKQQLFAEASGARGPVKVLVWTDGSSDPV 151

```

```

Db      498 TKVEDITIAINTFPYRGSSTNTGKMTYREKIFVPSKGSNNVPRKMITLTDGSSDAF 557
Qy      152 GPMQELKDLGVTVIVTSGRGNFLELSAASAPAEKHLHFV-DVDDHIIYOELRGS- 209
Db      558 RDPALIKRNSDVEIRAVGVKAQVRSLELAISPPAFETHVFTVEDPDAFORISFELTQISIC 617
Qy      210 -----LDAMR-----POQLHATEITSSGFLAMPPL-----LTADSGYIV 244
Db      618 LRIEQLAIAKKKAVVPPKDLFSFSEVTSYGFKTNMSPAGENVSFHYITKEAAGDEYIV 677
Qy      245 LE-----LVPSAQP-----GAARQQLPQNATD 267
Db      678 VEPASTSVLSLKPETILYLVNVTAEYEDGSIPLAGEETTEEKGAAPRLKATDETDT 737
Qy      268 -----WMA-----GLDPTDYLVALVPSN 288
Db      738 SFKITWQAQGRVLRCLIIYRPVAGSEREVTPPNORRTLENLIPPTKVEVSIVPEVF 797
Qy      289 VRLARPQLLRVTRPEEAGPERIVISHARPSRLSVNAPALGSAAALGYHVOGFPRLGCE 348
Db      798 SGPGTPTLGNMATEBEVRGNPRDLRVSDPTSTMKLSWGAQAPGKVKQ--YLVTYTPVAGGE 855
Qy      349 AQRVEVPAGRNCTTLOGIAPGTAVLVVTAAFRSGRESALSAAKACT 394
Db      856 TOEVTVRGDTTNTVLOGLKEGTQYALSVTALYASGADLFGEGIT 901

```

```

RESULT 3
US-11-169-041-160
; Sequence 160, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 160
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-169-041-160

```

```

Query Match      18.6%; Score 401; DB 7; Length 517;
Best Local Similarity 30.0%; Pred. No. 2,2e-24;
Matches 114; Conservative 61; Mismatches 179; Indels 26; Gaps 6;
Qy      36 MFLDLSASVSHYEFSSRVREYQGLVAPLPLGTALRASLVHVGSRPYTEPFQGHSSGE 95
Db      143 MFLVDGSGYIGIANFVKRAFLFVVKSFELSPNRRVOISLVOYSRDPTHEFLTKFTYE 202
Qy      96 AAQADVRAQAQMGDTHGLALVYAKQQLFAEASGARGPVKVLVWTDGSSDPVGPMP 155
Db      203 DIIEAINTFPYRGSSTNTGKMTYREKIFVPSKGSNNVPRKMITLTDGSSDAFRDPA 262
Qy      156 QELKDLGVTVIVTSGRGNFLELSAASAPAEKHLHFV-DVDDHIIYOELRGS- 209
Db      263 IKLRNSDVEIRAVGVKAQVRSLELAISPPAFETHVFTVEDPDAFORISFELTQISIC 322
Qy      210 -----LDAMR-----POQLHATEITSSGFLAMPPLITADSGYIVLEVPSAQAARQQL 261
Db      323 QELAIKKKAVVPPKDLFSFSEVTSYGFKTNMSPAGENVSFHYI-----TYKEAAGDEYV 376
Qy      262 -----PGNATDWMAGLDPTDYLVALVPSNRLRLPOLLRVTRPEEAGPERIVISHAR 317
Db      377 TVVEPASSTSVLNSLKPEETILYLVNVTAEYEDGFSIPLAGEETTEEKGAAPRLKATDET 436

```



```
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCN3
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-192-449-6
```

```
Query Match          9.0%; Score 194; DB 7; Length 214;
Best Local Similarity 33.5%; Pred. No. 1,6e-08;
Matches 63; Conservative 28; Mismatches 77; Indels 20; Gaps 7;
```

```
QY 34 DLMFLDSSASVSHYEFSEVREVFQGLVAPLPLGTGALRASLVHVSPPYTFPPGQHS 93
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 22 DIVIVLDGSNSI--YPMDSVTAFLNDLKRMDIGKQGVIGVGENVTHEFNINKYSS 79
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 94 GEAADAVASAGRMG-DHTGTLALVYAKEQLFAASGARPGVPVLYVWTTGGSSD--P 150
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 80 TEEVLVAARKIVQRGQGLQMTALGDTTARKEAFTARGARRGVKVMVITVDSHNDNR 139
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 151 VGPPMOELKDLGVTFIV-----STGRGNFL-----ELSAASAPAEKHLHFVDVDD-- 197
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 140 LKQVIODCDENIQRFSAIILGSYNRGNLSTKFEVEIKSIASEPTK--HFFNVSDLA 197
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 198 LHIIVQEL 205
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 198 LVTIVKTL 205
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

```
RESULT 7
US-11-192-449-9
; Sequence 9, Application US/11192449
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCN3
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
```

```
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-192-449-9
```

```
Query Match          9.0%; Score 194; DB 7; Length 214;
Best Local Similarity 33.5%; Pred. No. 1,6e-08;
Matches 63; Conservative 28; Mismatches 77; Indels 20; Gaps 7;
```

```
QY 34 DLMFLDSSASVSHYEFSEVREVFQGLVAPLPLGTGALRASLVHVSPPYTFPPGQHS 93
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 22 DIVIVLDGSNSI--YPMDSVTAFLNDLKRMDIGKQGVIGVGENVTHEFNINKYSS 79
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 94 GEAADAVASAGRMG-DHTGTLALVYAKEQLFAASGARPGVPVLYVWTTGGSSD--P 150
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 80 TEEVLVAARKIVQRGQGLQMTALGDTTARKEAFTARGARRGVKVMVITVDSHNDNR 139
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 151 VGPPMOELKDLGVTFIV-----STGRGNFL-----ELSAASAPAEKHLHFVDVDD-- 197
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 140 LKQVIODCDENIQRFSAIILGSYNRGNLSTKFEVEIKSIASEPTK--HFFNVSDLA 197
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 198 LHIIVQEL 205
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 198 LVTIVKTL 205
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

```
RESULT 8
US-11-192-449-5
; Sequence 5, Application US/11192449
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCN3
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Rat
US-11-192-449-5
```

```
Query Match          8.6%; Score 185; DB 7; Length 214;
Best Local Similarity 31.9%; Pred. No. 8e-08;
Matches 60; Conservative 32; Mismatches 76; Indels 20; Gaps 7;
```

```
QY 34 DLMFLDSSASVSHYEFSEVREVFQGLVAPLPLGTGALRASLVHVSPPYTFPPGQHS 93
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 22 DIVIVLDGSNSI--YPMDSVTAFLNDLKRMDIGKQGVIGVGENVTHEFNINKYSS 79
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 94 GEAADAVASAGRMG-DHTGTLALVYAKEQLFAASGARPGVPVLYVWTTGGSSD--P 150
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 80 TEEVLVAARKIVQRGQGLQMTALGDTTARKEAFTARGARRGVKVMVITVDSHNDNR 139
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 151 VGPPMOELKDLGVTFIVST-----GRGNFL-----ELSAASAPAEKHLHFVDVDD-- 197
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 140 LKQVIODCDENIQRFSAIILGSYNRGNLSTKFEVEIKSIASEPTK--HFFNVSDLA 197
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 198 LHIIVQEL 205
   |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

Db 198 LVTIVRAL 205

RESULT 9

US-10-063-703-34

Sequence 34, Application US/10063703

Publication No. US20060008901A1

GENERAL INFORMATION:

APPLICANT: Eaton, Dan L.

APPLICANT: Filvaroff, Ellen

APPLICANT: Geritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3230R1C1

CURRENT APPLICATION NUMBER: US/10/063,703

Prior Filing Date: 2002-05-08

NUMBER OF SEQ ID NOS: 170

SEQ ID NO 34

LENGTH: 678

TYPE: PRT

ORGANISM: Homo Sapien

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-10-063-703-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Prior Application Number: 60/170262

Prior Filing Date: 199-12-09

NUMBER OF SEQ ID NOS: 170

SEQ ID NO 34

LENGTH: 678

TYPE: PRT

ORGANISM: Homo Sapien

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;

Db 198 LVTIVRAL 205

US-11-102-240-34

Query Match

Best Local Similarity 29.9%; Pred. No. 4,4e-07;

Matches 55; Conservative 26; Mismatches 99; Indels 4; Gaps 2;


```

, PRIOR FILING DATE: 2001-03-19
, PRIOR APPLICATION NUMBER: 60/199476
, PRIOR FILING DATE: 2000-03-25
, PRIOR APPLICATION NUMBER: 09/863776
, PRIOR FILING DATE: 2001-05-23
, PRIOR APPLICATION NUMBER: 60/208263
, PRIOR FILING DATE: 2000-05-31
, PRIOR APPLICATION NUMBER: 09/93398
, PRIOR FILING DATE: 2001-08-24
, PRIOR APPLICATION NUMBER: 60/227800
, PRIOR FILING DATE: 2000-08-25.
, Remaining Prior Application data removed - See File Wrapper or PALM.
, NUMBER OF SEO ID NOS: 1609
, SOFTWARE: Curaseqslst version 0.1
, SEO ID NO 182
, LENGTH: 709
, TYPE: PRT
, ORGANISM: Homo sapiens
US-10-453-372-182

```

Query Match	8.0%;	Score 173;	DB 6;	Length 709;
Best Local Similarity	29.4%;	Pred. NO. 3.5e-06;		
Matches 68;	Conservative 33;	Mismatches 96;	Indels 34;	Gaps 8;

```

Oy      7 LGIATSLRATLARGSGERPPASAPRGDMLFLUDSSASVSHYEFSSRAREVQGLVAPL 66
Db      49 LGOARRRRRLRLRELSE-----RLVLFLVDDSSVGEVUNFRSELMFVRKLISFPV 100

Oy      67 GTGALRASLVHVGSRPYTEFPGOHSSGEAAQ-----DAVRASQRMGDTHGTALVY 119
Db      101 VPTARRVALVTFSSKNVYVPRVDYISTRARQHKKALLLOEIPALISYRGCGYTTKAFQ 166

Oy      120 AKEQLPAEASGARPGVVKYLVWVTTDGGSS--DPVGGPMOELKDLGVTVFVSTGRGNFLE 177
Db      161 AAQIILL-----HARENSTKVVSILITDYSNGDP-RLIASLRDSGVEIFFGIQGNIRE 215

Oy      178 LSAASAPAEKLIHFVDVDDLIHYVGLKGSILIDARPOQLVATETLTSSTF 228
Db      216 LNDMSTPEEHCYL-----LHSFEE-----FEALARRLIH--EDLPSSGF 254

```

RESULT 15
 US-10-601-368-18
 Sequence 18, Application US/10601368
 Publication No. US20050260702A1
 GENERAL INFORMATION:
 APPLICANT: Pan, Yang
 APPLICANT: Lora, Jose M.
 TITLE OF INVENTION: NOVEL INTEGRIN ALPHA SUBUNIT AND USES THEREOF
 FILE REFERENCE: 07334-275001
 CURRENT APPLICATION NUMBER: US/10/601,368
 CURRENT FILING DATE: 2003-06-23
 PRIOR APPLICATION NUMBER: US/09/561,263A
 PRIOR FILING DATE: 2000-04-27
 PRIOR APPLICATION NUMBER: US 09/322,750
 PRIOR FILING DATE: 1999-05-28
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 18
 LENGTH: 1167
 TYPE: PRN
 ORGANISM: Homo sapiens
 US-10-601-368-18

Query Match	8.0%;	Score 172.5;	DB 6;	Length 1167;
Best Local Similarity	24.0%;	Pred.No. 7.3e-06;		
Matches 90;	Conservative 47;	Mismatches 119;	Indels 119;	Gaps 16;

QY 18 ARSGAEKGPAP-APRG-----DLMFLDDSSASVSHYEFKSVREVGQVAPRLPGTGC 69
 QY ||| ||| ::||| ::||| ::||| :
 Db 143 ARVDASFGQGSLLAPRQRCPTTMDVYIVLDGSNSI--YPMSEVQITPLRLVKGKLPIDPE 200
 QY ||| ||| ::||| ::||| ::||| :
 QY 70 ALRALVNHGSRPYTEFFPGQSSSGEAAQADVAPASQRMG-DHTGTALVYAEQFLPAEA 128

Db 201 QIQVGLVYGSESPVHEWMSLGDPRTEKEEVYRAAKNLSRRREGREKTKQALIMVACTEGSSQ 266
QY 129 SGARPGVKVLVWTTDGGSDPVGPPMOEELKDLGTVTFVVSGR-----GNFL-- 176
Db 261 HGRPEEARLLVWVTDGESHDS-----EELPALKACEGRTYRYGIAVLGHYLRQ 312
QY 177 -----ELSAASAAPAKHLHFVDVDLHIIVGELRGSIIIDAMP-----QQLATAEI 223
Db 313 RDPSSFLREIKRTIADPDGR--FFNNVTD-----EALATDIVALGDRIFGLEGSHANE 365
QY 224 TSSGGRLLAMPPLLTADSGYVLELVPSAQPAGARRQQLPG-----NATDW-----IWAGL 273
Db 366 SSFGLGM-----SQIGSTRHLKDGILFCMGVAGIDMGQSVLML-- 403
QY 274 DPDTIDVALVPESVRLRLPOLRLVRRTRPEAGPERIVISHARPSLSKVSAPAL-GSA 332
Db 404 -----EGCHRLFPPRM-----ALDEDFPPALONHA 428
QY 333 AALGYHVOFGPLRG 347
Db 429 AYLIGYSVSSMLIRGG 443

Search completed: February 13, 2006, 13:12:59
Job time : 10.087 secs

THIS PAGE BLANK (USPTO)

Db 92 SSGEAAQAVRASAGRMEDTHTGLALVYAKEDLFEASGARGVVKLVWVTGGSSDPV 151
QY 121 GPPMOELDGLGTVIVTGTGKGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLD 180
Db 152 GPPMOELDGLGTVIVTGTGKGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLD 211

RESULT 2

US-11-186-284-26
Sequence 26, Application US/11186284
Publication No. US20050266493A1

GENERAL INFORMATION:

APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
APPLICANT: Burgart, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF COLON CANCER
FILE REFERENCE: MEM01-029F28RM
CURRENT APPLICATION NUMBER: US/11/186,284
PRIOR FILING DATE: 2005-07-21
PRIOR APPLICATION NUMBER: US/10/301,822
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 3063
TYPE: PRT
ORGANISM: Homo Sapiens
US-11-186-284-26

Query Match 28.9%; Score 263.5; DB 7; Length 3063;
Best Local Similarity 35.8%; Pred. No. 1.1e-17;
Matches 64; Conservative 30; Mismatches 84; Indels 1; Gaps 1;

QY 1 RGLDMFLDSSASVSHYEFRRVFGQLVAPLPLGTGALRASLVHVGSRPTTFEPFGH 60
Db 438 KADIVFLVDGSSISGIANFVKRAFLVLYKSFSPKRVQISLVQYSRDPTEFTLKKF 497
QY 61 SSGEAAQAVRASAGRMEDTHTGLALVYAKEDLFEASGARGVVKLVWVTGGSSDPV 120
Db 498 TKVEIITIEIINTFPFGSGSTNTGKAMTYREKIFVPSKGSRSNPKWMLITDGSDDAF 557
QY 121 GPPMOELDGLGTVIVTGTGKGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGSI 178
Db 558 RDPALKNSDVEIFAVGVKDAVRSLELAISPPAETHVFTVEDFDAPQRISEFELTQSI 616

RESULT 3

US-11-169-041-160
Sequence 160, Application US/11169041
Publication No. US20060019284A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
FILE REFERENCE: 10001 NP
CURRENT APPLICATION NUMBER: US/11/169,041
CURRENT FILING DATE: 2005-06-28

;; PRIOR APPLICATION NUMBER: 60/584,405
;; PRIOR FILING DATE: 2004-06-30
;; NUMBER OF SEQ ID NOS: 527
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 160
;; LENGTH: 517
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-11-169-041-160

Query Match 28.2%; Score 257.5; DB 7; Length 517;
Best Local Similarity 36.6%; Pred. No. 4.1e-18;
Matches 64; Conservative 27; Mismatches 83; Indels 1; Gaps 1;

QY 5 MFLDSSASVSHYEFRRVFGQLVAPLPLGTGALRASLVHVGSRPTTFEPFGHSSGE 64
Db 143 MFLVDGSSISGIANFVKRAFLVLYKSFSPKRVQISLVQYSRDPTEFTLKKFIVE 202
Db 203 DIIETINTFPFGSGSTNTGKAMTYREKIFVPSKGSRSNPKWMLITDGSDDAFRDP 262
QY 125 QELMDLGLGTVIVTGTGKGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGSI 178
Db 263 IKLRNSDVEIFAVGVKDAVRSLELAISPPAETHVFTVEDFDAPQRISEFELTQSI 317

RESULT 4

US-10-131-826A-294
Sequence 294, Application US/10131826A
Publication No. US20050245730A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gunney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C128
CURRENT APPLICATION NUMBER: US/10/131,826A
PRIOR FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588

```

; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 294
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-294

Query Match      27.5%; Score 251.5; DB 6; Length 915;
Best Local Similarity 32.2%; Pred. No. 3.5e-17;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

QY 1 RQDLFPLDSSASVSHYEFRRVEVQGVAPLPGTGLRASLVHSGRPYTEPPFGQH 60
DB 55 RADLVFIIDSSRSVNHDAKYAEFIVDILQFLDIGPDVTRVGLLQYGSSTVKNFSLKTF 114
QY 61 SSGEAAQDVAVRASQMGSTHGLALVYAKQQLFAEASGARP---GVPRVLVWMTDGGSS 117
DB 115 KRKSEVERAVKRWRLSTGTMTGLAQYALNTAFSSAEGARPLRENVPRVIMIVTDGRPO 174
QY 118 DPVGPMPQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHL---HFVDDDLHIYOE 173
DB 175 DSVAEVAAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVFOK 234

RESULT 5
US-11-113-424-39
; Sequence 39, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangoli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; PRIOR FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 39
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-39

Query Match      27.5%; Score 251.5; DB 7; Length 956;
Best Local Similarity 32.2%; Pred. No. 3.7e-17;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

QY 1 RQDLFPLDSSASVSHYEFRRVEVQGVAPLPGTGLRASLVHSGRPYTEPPFGQH 60
DB 55 RADLVFIIDSSRSVNHDAKYAEFIVDILQFLDIGPDVTRVGLLQYGSSTVKNFSLKTF 114
QY 61 SSGEAAQDVAVRASQMGSTHGLALVYAKQQLFAEASGARP---GVPRVLVWMTDGGSS 117
DB 115 KRKSEVERAVKRWRLSTGTMTGLAQYALNTAFSSAEGARPLRENVPRVIMIVTDGRPO 174
```

```

QY 118 DPVGPMPQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHL---HFVDDDLHIYOE 173
DB 175 DSVAEVAAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVFOK 234

RESULT 6
US-11-192-449-6
; Sequence 6, Application US/11192449
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Genzyme, Inc.
; APPLICANT: Koteliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USC93
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-192-449-6

Query Match      21.2%; Score 194; DB 7; Length 214;
Best Local Similarity 33.5%; Pred. No. 2.7e-12;
Matches 63; Conservative 28; Mismatches 77; Indels 20; Gaps 7;

QY 3 DLMFLDSSASVSHYEFRRVEVQGVAPLPGTGLRASLVHSGRPYTEPPFGQSS 62
DB 22 DIVIVLDGSSNSI--PFWQSVTAFPLNDLKRMDIGKQOVGVQYGENVTHEFNLNKXSS 79
QY 63 GEAAQDAVRASQMGSTHGLALVYAKQQLFAEASGARPQVPLVWMTDGGSSD--P 119
DB 80 TEEVVAAKKIVQGRGRTMTALGDTARKAFTLARKARRGVKVMYIVTDGESHDFR 139
QY 120 VGPMPQELKDLGVTVFIV---STGRGNFL-----ELSAASAPAEKHLHFVDDVDD--- 166
DB 140 LKQVQDCEDENIQFSAIIGSVYRGNLSTKQFVEIKSIASEPTEK--HFFNVSDELA 197
QY 167 LHITVQEL 174
DB 198 LVTIVKTL 205

RESULT 7
US-11-192-449-9
; Sequence 9, Application US/11192449
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Genzyme, Inc.
; APPLICANT: Koteliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USC93
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
```

```

; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-192-449-9
```

```
Query Match      21.2%; Score 194; DB 7; Length 214;
Best Local Similarity 33.5%; Pred. No. 2.7e-12;
Matches 63; Conservative 28; Mismatches 77; Indels 20; Gaps 7;
```

```
QY 3 DLMLDSSASVSHYEFRRVREFGQVAPLPLGTGALRASLVHVGSRPYTEFPFGQHS 62
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 22 DIVVLDSNSI--YPMDSVTAFLNDLKRMDIGPKQVGVIGENVTHEFNLKYS 79
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 63 GEAADAVRASQRMG-DTHTGLALVYAKEQLFAASGARPGVPLVWVTDDGSSD--P 119
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 80 TEEVLVAAKKIVQGRGQMTALGDTYAKFAFGARGVKKVWVIVTDGSHDNMR 139
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 120 VGPMQELKDGVTVFIV---STGRGNFL-----ELSAASAPAEKHLHFVVDVDD-- 166
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 140 LKQVIQDQDENIQFSAIILGSYVRGNLSTKEFVEIKSIASEPTEK--HFNVSDELA 197
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 167 LHIIVQEL 174
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 198 LVTIVKTL 205
```

```

RESULT 8
US-11-192-449-5
; Sequence 5, Application US/11192449
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Koceliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCN3
; CURRENT APPLICATION NUMBER: US/11/192,449
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Rat
US-11-192-449-5
```

```
Query Match      20.3%; Score 185; DB 7; Length 214;
Best Local Similarity 31.9%; Pred. No. 2.1e-11;
Matches 60; Conservative 32; Mismatches 76; Indels 20; Gaps 7;
```

```
QY 3 DLMLDSSASVSHYEFRRVREFGQVAPLPLGTGALRASLVHVGSRPYTEFPFGQHS 62
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 22 DIVVLDSNSI--YPMDSVTAFLNDLKRMDIGPKQVGVIGENVTHEFNLKYS 79
```

```
QY 63 GEAADAVRASQRMG-DTHTGLALVYAKEQLFAASGARPGVPLVWVTDDGSSD--P 119
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 80 TEEVLVAAKKIVQGRGQMTALGDTYAKFAFGARGVKKVWVIVTDGSHDNMR 139
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 120 VGPMQELKDGVTVFIVST---GRGNFL-----ELSAASAPAEKHLHFVVDVDD-- 166
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 140 LKQVIQDQDENIQFSAIILGSYVRGNLSTKEFVEIKSIASEPTEK--HFNVSDELA 197
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 167 LHIIVQEL 174
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 198 LVTIVKAL 205
```

```

RESULT 9
US-10-063-703-34
; Sequence 34, Application US/10063703
; Publication No. US2006008901A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Macanabe, Colin K.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,703
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 34
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-703-34
```

```
Query Match      20.0%; Score 183; DB 6; Length 678;
Best Local Similarity 30.6%; Pred. No. 1.6e-10;
Matches 53; Conservative 25; Mismatches 91; Indels 4; Gaps 2;
```

```
QY 3 DLMLDSSASVSHYEFRRVREFGQVAPLPLGTGALRASLVHVGSRPYTEFPFGQHS 62
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 495 DIGFVIDGSSVGVGNRTYVQFTNLTKEFEISDTRIGAVGYTGEORLEPFQDKXS 554
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 63 GEAADAVRASQRMGDTHTGLALVYAKEQLFAASGARPGVPLVWVTDDGSSDPVGP 122
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 555 KPDIINAIKRVYSGGTSGAINFALQELFKK---SKPKRKLMILITDGRSYDVRI 611
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 123 PMQELKDGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDVDDV 174
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 612 PAMAHLKGVITVYAIIVAMAQEELEVIATHPADHSFFVEDEFNLHQVYRI 664
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```

RESULT 10
US-11-102-240-34
; Sequence 34, Application US/11102240
; Publication No. US20050260647A1
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: ANTIBODIES TO POLYPEPTIDES ENCODED BY A NUCLEIC ACID UNDEREXPRESS
; FILE REFERENCE: P3230R1C106C
; CURRENT APPLICATION NUMBER: US/11/102,240
; CURRENT FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: 10/063662
```

```

; PRIOR FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: 10/006867
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/170262
; PRIOR FILING DATE: 199-12-09
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 34
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-102-240-34

```

```

Query Match      20.0%; Score 183; DB 7; Length 678;
Best Local Similarity 30.6%; Pred. No. 1.6e-10;
Matches 53; Conservative 25; Mismatches 91; Indels 4; Gaps 2;

```

```

QY 3 DLMFLDSSASVSHYEFRRVEFGQLVAPLPLGTGALRASLVHVGSRPYTEPFQGHSS 62
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 495 DIGFVIDSSSVGTGNFRVLQFVTNLKFEISDTRIGAVQYTYEQRLFEFGFDKYS 554
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 GEAADAVRASAQRMGDTHTGLALVYAKQLFAASGARGVPGKVLWVTDGSSSDPYG 122
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 555 KPDIILAIKRVGWSGISTGAALINFLDELFPK--SKPKKKMLITDGRSYDVR 611
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 PMOELKDLGVTVFVSTGRGNFLSAAASAPAEKHLHFVD-VDDLHIYQEL 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 612 PMAAHLKGVITYTIGVAMAAGELEVIATNHPARDHSFFDFEDNHLQVYPR 664
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-10-453-372-186
; Sequence 186, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 186
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-186

```

```

Query Match      17.9%; Score 163; DB 6; Length 709;
Best Local Similarity 31.4%; Pred. No. 1.7e-08;
Matches 54; Conservative 27; Mismatches 75; Indels 16; Gaps 5;

```

```

QY 1 RGDIMFLDSSASVSHYEFRRVEFGQLVAPLPLGTGALRASLVHVGSRPYTEPFQGH 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 RELVFLVDDSSVGVNRSRLMFVKLLSDPFVPTATRAIVTFSSKNVY-VRADY 124
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 SSGEAAQ-----DAVASAQRMGDTHTGLALVYAKQLFAASGARGVPGKVLWV 112
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 ISTRARAROKCALLLOEIPALISYRGCGTYTKGAFQQAQIIL---HARENSAKVFLIT 180
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 DGGSS--DPVGPMPQELKDLGVTVFVSTGRGNFLSAAASAPAEKHLHFV 162
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 181 DGYSGDGF-RPLAASLRDSGVEITFGIWOGNRELNMMASTPKKEHCYLL 231
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
US-11-080-026-4
; Sequence 4, Application US/11080026
; Publication No. US20050260192A1
; GENERAL INFORMATION:
; APPLICANT: Springer, Timothy A.
; APPLICANT: Shimooka, Motomu
; APPLICANT: Lu, Chafan
; TITLE OF INVENTION: MODIFIED POLYPEPTIDES STABILIZED IN A
; FILE REFERENCE: CFEP-P02-021
; CURRENT APPLICATION NUMBER: US/11/080,026
; PRIOR FILING DATE: 2005-03-15
; PRIOR APPLICATION NUMBER: 09/945,265
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,700
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1152
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-080-026-4

```

```

Query Match      17.8%; Score 162.5; DB 7; Length 1152;
Best Local Similarity 28.1%; Pred. No. 3.6e-08;
Matches 52; Conservative 31; Mismatches 91; Indels 11; Gaps 5;
QY 3 DLMFLDSSASVSHYEFRRVEFGQLVAPLPLGTGALRASLVHVGSRPYTEPFQGHSS 62
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 150 DIAFLDSSGSIIPDFRMRKFFVSTVMEQLKSKTLF--SLMQSSEERHIFTPEFQ 207
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 GEAADAVRASAQRMGDTHTGLALVYAKQLFAASGARGVPGKVLWVTDGSS-SDPYG 121
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 208 NNPRSLVPIQTGLRFTATGIRKRVARELFINTGARKNAFKILVITDEKFKFDPLG 267
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 --PMOELKDLGVTVFVSTG-----RGNFLELSAAASAPAEKHLHFV-DVDDLHIYQEL 173
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 266 YEDVLPEDAREGVIRVIGVGDAFRSEKSRDELNTIASKPPRDHVFQVNNFALKTIQ 327
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 LRGSI 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 LREKI 332
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-10-453-372-180
; Sequence 180, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967

```

```

RESULT 14
US-10-453-372-194
; Sequence 194, Application US/10453372
; Publication No. US20060003323a1
; GENERAL INFORMATION:
; APPLICANT: Alsdbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25

```

```

    / PRIOR APPLICATION NUMBER: 09/789390
    / PRIOR FILING DATE: 2001-02-23
    / PRIOR APPLICATION NUMBER: 60/185967
    / PRIOR FILING DATE: 2000-03-01
    / PRIOR APPLICATION NUMBER: 09/823187
    / PRIOR FILING DATE: 2001-03-29
    / PRIOR APPLICATION NUMBER: 60/195792
    / PRIOR FILING DATE: 2000-03-10
    / PRIOR APPLICATION NUMBER: 09/839446
    / PRIOR FILING DATE: 2001-03-19
    / PRIOR APPLICATION NUMBER: 60/199476
    / PRIOR FILING DATE: 2000-03-25
    / PRIOR APPLICATION NUMBER: 09/863776
    / PRIOR FILING DATE: 2001-05-23
    / PRIOR APPLICATION NUMBER: 60/208263
    / PRIOR FILING DATE: 2000-05-31
    / PRIOR APPLICATION NUMBER: 09/939398
    / PRIOR FILING DATE: 2001-08-24
    / PRIOR APPLICATION NUMBER: 60/227800
    / PRIOR FILING DATE: 2000-08-25
    / Remaining Prior Application data removed - See File Wrapper or PALM.
    / NUMBER OF SEQ ID NOS: 1609
    / SOFTWARE: Curaseq1st version 0.1
    / SEQ ID NO 178
    / LENGTH: 3570
    / TYPE: PRT
    / ORGANISM: Homo sapiens
US-10-453-372-178

```

Qy 1 RGDIMFLDDSSASVSHYEFSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYEEFPGQH 60
:
:
:
Dd 78 RLIEFLVYDDSSSGVEINFRSELMFVRKLSDFEVPPTTRVALVFSSKNVYPVRVDYI 137
:
:
:
:

OY 61 SSGEAAQ-----DAVRAAQRMGDTHTGLALVYAKEOLFPAASGARPGVKULWVTD 113
 Db 138 STRRARQHKCALLLOEIPALSYRGGTYTKGAFQQAQIL-----HARENSTKVVFLITD 193
 OY 114 GGSS--DVGPPMDELKDLGVTVFIVSTGRGNFELSAAASAPAEKHLHFV 162
 Db 194 GYSNGGDP-RPIAASLRDSGVEIFTFGIWQGNIRELNDMASTPKKEHCYTL 243

Search completed: February 13, 2006, 13:12:58
 Job time : 3.91304 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:44:15 ; Search time 2.24054 Seconds
(without alignments)
1468.904 Million cell updates/sec

Title: US-10-699-035A-5
Perfect score: 2380
Sequence: 1 atgtccctccctgagcgcgcct.....ccgcacgcctgagcgcgttaa 1254

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Fgapop 10.0, Fgapext 0.5
Delop 6.0, Delext 7.0

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-O=/abs/ABSGWEE spool/US10699035/runat_13022006_062453_25634/app_query.fasta_1
-DB=Published Applications_AA_New -QPMT=faetan -SUFFIX=rapbn -MINMATCH=0.1
-DOOPL=0 -DOOPEXT=0 -UNITS=bits -STAR1=1 -END=-1 -MATRIX=bl0sum62
-TRANS=human40.cdi -LIST=45 -LOCAL=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abs82p
-USER=US10699035 @CGN 1 1 10 @runat_13022006_062453_25634 -NCP1=6 -ICPU=3
-NO_MMAPP -NBS_SCORES=0 -WAIT_DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA_New:*
1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2114	88.8	445	6	US-10-453-372-2
2	419	17.6	3063	7	US-11-186-284-26
3	398.5	16.7	517	7	US-11-169-041-160
4	254.5	10.7	1717	7	US-11-182-016-20
5	253.5	10.7	915	6	US-10-131-826A-294
6	221.5	9.3	1366	6	US-11-113-424-39
7	221.5	9.3	1366	6	US-10-821-234-1431
8	221.5	9.3	1366	7	US-11-186-284-31
9	215	9.0	1496	7	US-11-186-284-35

10	212	8.9	1464	7	US-11-000-463-243	Sequence 243, App
11	212	8.9	1464	7	US-11-186-284-28	Sequence 28, Appl
12	212	8.9	1464	7	US-11-021-603-2	Sequence 2, Appl1
13	212	8.9	1467	6	US-10-821-234-1096	Sequence 1096, Ap
14	211	8.9	1464	7	US-11-000-463-243	Sequence 243, App
15	211	8.9	1464	7	US-11-186-284-28	Sequence 28, Appl
16	211	8.9	1464	7	US-11-021-603-2	Sequence 2, Appl1
17	211	8.9	1467	6	US-10-821-234-1096	Sequence 1096, Ap
18	210.5	8.8	1466	6	US-11-821-234-964	Sequence 964, App
19	210.5	8.8	1466	6	US-11-186-284-33	Sequence 33, Appl
20	206.5	8.7	1733	7	US-11-182-016-21	Sequence 21, Appl
21	203.5	8.5	1466	7	US-11-186-284-33	Sequence 33, Appl
22	202	8.5	1733	7	US-11-182-016-21	Sequence 21, Appl
23	200.5	8.4	1823	6	US-10-995-561-998	Sequence 998, App
24	200.5	8.4	2102	6	US-10-995-561-990	Sequence 990, App
25	200.5	8.4	2108	6	US-10-995-561-989	Sequence 989, App
26	200.5	8.4	2157	6	US-10-995-561-991	Sequence 991, App
27	199	8.4	214	7	US-11-192-449-6	Sequence 6, Appl1
28	199	8.4	214	7	US-11-192-449-9	Sequence 9, Appl1
29	199	8.4	1717	7	US-11-182-016-20	Sequence 20, Appl
30	198.5	8.3	1366	6	US-10-821-234-1431	Sequence 1431, Ap
31	198.5	8.3	1366	6	US-11-186-284-31	Sequence 31, Appl
32	198.5	8.3	1742	7	US-11-182-016-23	Sequence 23, Appl
33	197	8.3	843	7	US-11-129-104-89	Sequence 89, Appl
34	196.5	8.3	3570	6	US-10-453-372-196	Sequence 196, App
35	195.5	8.2	3570	6	US-10-453-372-178	Sequence 178, App
36	195.5	8.2	3570	6	US-10-453-372-198	Sequence 198, App
37	195.5	8.2	3570	6	US-10-453-372-200	Sequence 200, App
38	195.5	8.2	3570	6	US-10-453-372-202	Sequence 202, App
39	195.5	8.2	3570	6	US-10-453-372-204	Sequence 204, App
40	195.5	8.2	3570	6	US-10-453-372-206	Sequence 206, App
41	192.5	8.1	1251	7	US-11-149-003-16	Sequence 16, Appl
42	192	8.1	709	6	US-10-453-372-182	Sequence 182, Appl
43	191	8.0	495	7	US-11-182-016-31	Sequence 31, Appl
44	190.5	8.0	709	6	US-10-453-372-180	Sequence 180, App
45	190.5	8.0	709	6	US-10-453-372-186	Sequence 186, App

ALIGNMENTS

RESULT 1
US-10-453-372-2
; Sequence 2, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsebrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHO
; FILE REFERENCE: 21402-589 A US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraSeqst version 0.1

Percent Similarity: 39.9%
Best Local Similarity: 27.5%
Query Match: 17.6%
DB: 7
Conservative: 58
Matches: 174
Indels: 106
Gaps: 9

US-10-699-035a-5 (1-1254) x US-11-166-284-26 (1-3063)

```
QY 94 CGAGGGAGCCTGATGTTCTGCTGACAGCTCAGCCAGCGCTCTCTCACTACGAGTTCTCC 153
DB 438 LysAlaAspIleValPheLeuValAspGlySerTyrSerIleGlyIleAlaAspPheVal 457
QY 154 CGGGTTCGGAGAGTTTGGGGCAGCTGGGCTCCACTGCGCCCTGGGGACCGGGGCGCTG 213
DB 458 LysValAlaArgAlaPheLeuGluValLeuValLysSerPheGluIleSerProAlaArgVal 477
QY 214 CGTCAGCTGCTGGTGCACGTGGGCACTGCGCATACAGGATTCCCTTCCGCGCAGCAG 273
DB 478 GlnIleSerLeuValGlnTyrSerArgAspProIleThrGlnPheThrLeuLysLysPhe 497
QY 274 AGCTCGGGTGAAGCTGCGCCAGATGCGGTGCGCTTCTGCGCCAGCGCATGGGTGACCC 333
DB 498 ThrLysValAlaGluAspIleIleGluAlaIleAsnThrPheProTyrArgLysSerThr 517
QY 334 CACACTGGCTGGCGCTGCTCTATGCCAAGAACACTGTTTGTGAAGCATCAGTGC 393
DB 518 AsnThrGlyLysAlaMetThrTyrValArgGluLysIlePheValProSerLysGlySer 537
QY 394 CGGCGAGGGGTGCGCCAAAGTGTGCTGGTGAAGATGATGCGGCGCTCCCGACCGCTG 453
DB 538 ArgSerAsnValProLysValMetIleLeuIleThrAspLysSerSerAspAlaPhe 557
QY 454 GGGCCCCCATGAGAGCTCAAGACCTGGGCGTCACTGCTTCTTCACTTGTACAGCCGCG 513
DB 558 ArgAspProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValLys 577
QY 514 CGAGGCACTCTCTGAGCTGTGACGCCGCTCAGCCCTCAGCCCTCCGAGAAGCACTGCAC 573
DB 578 AspAlaValArgSerGluLeuGluAlaIleAlaSerProProAlaGluThrHisValPhe 597
QY 574 TTGTG---GACGTGATGACCTGCACTATGTCAGAGAGTGGGGCTCATTT--- 627
DB 598 ThrValGluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCys 617
QY 628 -----CTGCGC-----ATGCGCGCGCAGCAG 648
DB 618 LeuArgIleGluGlnGluLeuAlaAlaIleLysLysLysAlaTyrValProProLysAsp 637
QY 649 CTCCTATGCCAGGAGATCAGCTCAGCGCTTCCGCTGCGCTGCGCACCCCTG----- 702
DB 638 LeuSerPheSerGluValThrSerTyrGlyPheLysThrAsnTrpSerProAlaGlyGlu 657
QY 703 -----CTGACCGGAGACTCGGGCTACTATG 729
DB 658 AsnValPheSerTyrHisIleThrTyrLysGluAlaAlaGlyAspAspGluValThrVal 677
QY 730 CTGAG-----CTGTGCGCGCAGCGCGCGCAGCG----- 756
DB 678 ValGluProAlaSerSerThrSerValValLeuSerSerLeuLysProGluThrLeuTyr 697
QY 756 ----- 756
DB 698 LeuValAsnValThrAlaGluTyrGluAspGlyPheSerIleProLeuAlaGlyGluGlu 717
QY 757 -----GGGCTGCAAGACCGCACAGCTGCGCAGGAAACCGCAGGAG 798
DB 718 ThrThrGluGluValLysGlyAlaProArgAsnLeuLysValThrAspGluThrThrAsp 737
QY 799 -----TGCATCTGGGCC----- 810
DB 738 SerPheLysIleThrTrpThrGlnAlaProGlyArgValLeuArgCysArgIleIleTyr 757
QY 810 ----- 810
```

```
DB 758 ArgProValAlaGlyLysSerArgLysValThrThrProProAsnGlnArgArgArg 777
QY 811 -----GGCCTCGACCCGAGCAGCAGCTACGAGTGGCGCTAGTGCCTGACCAAC 861
DB 778 ThrLeuGluAsnLeuIleProAspThrLysTyrGluValSerValIleProGluTyrPhe 797
QY 862 GTGCGCTCTTGAAGCCCAAGATCCTGCGGGTCCGACCGCGCAGAGAGCCGGGCCA 921
DB 798 SerGlyProGlyThrProLeuThrArgLysAsnAlaAlaThrGluGluValArgLysAsnPro 817
QY 922 GAGCGCATGCTATCTCCACCGCGCGCGCGCGCGCTCGGCTGAGTGGGCGCCAGC 981
DB 818 ArgAspLeuArgValSerAspProThrThrSerThrMetLysLeuSerTrpSerGlyAla 837
QY 982 CTGGCTCAGCCGCGCGCTCGGCTACCACTGCACTGAGTGGCGCGCTGCGCGCGGAG 1041
DB 838 ProGlyLysValLysGln-----TyrLeuValThrTyrThrProValAlaGlyGlyGlu 855
QY 1042 GCGCAGCGGGTGAAGTGCCTCCGCGCGCGCGCACTGACCACTGACAGGCGCTGGCGCC 1101
DB 856 ThrGlnGluValThrValArgLysAspThrThrAsnThrValLeuGlnGlyLeuLysGlu 875
QY 1102 GGCACCGCTTACCTGCTGACCGTGAACCGCGCTTCCGCTCGGGCGCGAGCGGCTG 1161
DB 876 GlyThrGlnTyrAlaLeuSerValThrAlaLeuTyrAlaSerGlyAlaGlyAspAlaLeu 895
QY 1162 TCCGCCAGAGCCTGCAGC 1179
DB 896 PheGlyGluGlyThrThr 901
```

RESULT 3

US-11-169-041-160
; Sequence 160, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 160
; LENGTH: 517
; TYPE: PRF
; ORGANISM: Homo sapiens
US-11-169-041-160

Alignment Scores:

Pred. No.: 2,83e-16 Length: 517
Score: 398.50 Matches: 114
Percent Similarity: 45.8% Conservative: 60
Best Local Similarity: 30.0% Indels: 179
Query Match: 16.7% Gaps: 27
DB: 7

US-10-699-035a-5 (1-1254) x US-11-169-041-160 (1-517)

```
QY 106 ARGTTCTGCTGAGACAGCTCAGCCAGCGTCTCTCACTACGAGTTCTCCGGGTTGGGAG 165
DB 143 MetPheLeuValAspGlySerTyrSerIleGlyIleAlaAsnPheValLysValArgAla 162
QY 166 TTGTGCGGCGACGTGGTGCCTCACTGCGCTGCGGAGCAGCGGAGCCCTGCGTGCAGTCTG 225
DB 163 PheLeuGluValLeuValLysSerPheGluIleSerProAsnArgValGlnIleSerLeu 182
QY 226 GTGCACGTGGGAGTGGCCATACACCGAGTTCCCTTCCGCGCAGCAGCTCGGGTGA 285
```

Db 183 ValGlnTYrSerArgAspProHisThrGluPheThrLeuLeuYLeuPheThrIlyValGlu 202
QY 286 GCTGCCAGATGCGATGCGTCTTCGCCAGCGATGAGTGAACCCACATGCGCTG 345
Db 203 AspiIleIleGluAlaIleLeuThrPheProTYrArgGluYGLYSerThrThrGlyLys 222
QY 346 GCGCTGCTATATCCAGAGACAGCTGTTGCTGAAGCATCGAGGCGCCAGGGG 405
Db 223 AlaMetThrTYrValArgGluYLeuIlePheValProSerIlySerArgSerAsnVal 242
QY 406 CCCAAGTCTGCTGTGGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCCCCCGC 465
Db 243 ProIlyValMetIleIleuIleThrAspGlyLysSerSerAspAlaPheArgAspProAla 262
QY 466 CAGAGCTTCAAGACCTGGCGCTCACCGTTCATTTGTCAGACCGCGCGCAACTTC 525
Db 263 IleIlyLeuArgAsnSerAspValGluIlePheAlaValGlyValIlyAspAlaValAsp 282
QY 526 CTGAGAGCTGTCAGCGGCTTCAGCGCTCCCGGAGAACACTGACATTTGTG--GAC 582
Db 283 SerGluLeuGluAlaIleAlaSerProAlaGluThrIleValPheThrValGluAsp 302
QY 583 GTGATGACCTGCATCATTCATTCACAGAGCTGAGGGGCTCCATT----- 627
Db 303 PheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCysLeuArgIleGlu 322
QY 628 -----CTCGCG-----ATCGCGCGCAGACAGCTCCATGCGCAG 660
Db 323 GlnGluLeuAlaAlaIleLysLysAlaTYrValProIlyAspSerPheSer 342
QY 661 GAGATCAGCTCCAGGGGCTTCGCGCTGCGCTGCCACCCCTGCTGACCCCAACTCGGGC 720
Db 343 GluValIThrSerTYrGlyPheLysThrAsnTPSerProAlaGlyLysAsnValPheSer 362
QY 721 TACTATGCTGAGAGCTGCTGCCAGCGCGCGGCTGTCAAGACCGCAGCAGCTG 780
Db 363 TyrHisIle-----ThrTYrLysGluAlaAlaIlyAspArgGluVal 376
QY 781 -----CCAGGAAACCGCAGACTGATCTGGCGCGCTGCACCCGAGCAGC 828
Db 377 ThrValValGluProlAserSerThrSerValValLeuAsnSerLeuYsProlGluThr 396
QY 829 GACTACGACGTGGCGCTAGCTGCTGATCAACAGTGGCGCTCTGAGGCGCCAGATCTG 888
Db 397 LeuTYrLeuValAsnValThrAlaGluTYrGluAspGlyPheSerIleProLeuAlaGly 416
QY 889 CGGAGCGCAGCGCGCAGAGAGGCGCGCGCAGAGCGATGATCTCCACGCGCGG 948
Db 417 GluGluThrThrGluGluValLysGlyAlaProAlaGlnLeuYsValThrAspGluThr 436
QY 949 CCGCGCAGCGCTCGCGTGAATTGGGCCCGCAGCGCTGGGCTCAGCGCGCGCTCGGCTAC 1008
Db 437 ThrAspSerPheIlyLeuThrSerThrGlnAlaProGlyLys-----ValLeuArgTYr 454
QY 1009 CACGTGCACTGGCGCGCTGGCGCGCGCGCGCGCGCGCGCGAGTGCAGCGCGG 1068
Db 455 ArgIleIleTYrArgProValAlaGlyGlyGluSerArgGluValThrThrProAsn 474
QY 1069 CGCACTGCGACACGAGGAGCTGTGGCGCGCGCGCGCGCGCGCTTACCTGAGCTGAC 1128
Db 475 GlnAlaGArgArgThrLeuGluAsnLeuIleProAspThrIlyTYrGluValSerValIle 494
QY 1129 GCCGCGCTTCGCGCGCGCGCAGAGCGCGCTGTCCGCCAGAGCGCTGACGCGCGCAG 1188
Db 495 ProGluTYrPheSerGlyProGlyThrProLeuThrGlyAsnAlaAlaThrGluGluGly 514

RESULT 4
US-11-182-016-20
; Sequence 20, Application US/11182016
; Publication No. US20060019294A1
; GENERAL INFORMATION:
; APPLICANT: SUGEN, INC.
; TITLE OF INVENTION: TYROSINE KINASE SUBSTRATE (TKS) PROTEINS

; FILE REFERENCE: 038602/0102
; CURRENT APPLICATION NUMBER: US/11/182,016
; CURRENT FILING DATE: 2005-07-15
; PRIOR APPLICATION NUMBER: US/09/958,359
; PRIOR FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 1717
; TYPE: prt
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Grub
US-11-182-016-20

Alignment Scores:
Pred. No.: 3,29e-08 Length: 1717
Score: 254.50 Matches: 161
Percent Similarity: 31.3% Conservative: 23
Best Local Similarity: 27.4% Mismatches: 155
Query Match: 10.7% Indels: 250
DB: Gaps: 39

US-10-699-035A-5 (1-1254) x US-11-182-016-20 (1-1717)

QY 1233 CCCCGGGGTGCGGCG-----CGGGGACAGGGGCGTGGCGCGGCGCGCGCTC 1183
Db 725 ProArgCysAlaGlyLeuGlyTYrAlaGlyIleGlyAlaGlyLysSerArgAlaArg 744
QY 1182 GGGCGTGAGAGCGCTTGGCGGACAGCGCGCTCGCG----- 1147
Db 745 GlyArgGlyGlySerGlyAsnAlaGlnAlaThrAlaGluGlyAlaGlyArgSerProAla 764
QY 1146 ---GCCCGAGCGGAAGC---GCCGCTCAC-----GCTGAC 1117
Db 765 AspGlyThrAlaGluGlyTYrGlyLysIleProAspArgAlaAlaLeuHisSerGlnGln 784
QY 1116 CAGTAGGCGGTGCGCGCGCGCGCGCGCGCTGACGCTGTCAGGCTGCGCGCGCGCAGC 1057
Db 785 GlnAlaGlyGlyProArgProSerTYrThrValSerGly-----SerGlyArgGlyHis 802
QY 1056 CT--CCAGCCGCTGCGCTCCCGCGCGCGCGCGCGCGCGCACTGACAGCTGTCAGCGCAGCG 998
Db 803 SerProAlaCysAlaProLeuGlnProAlaArgAla---AlaAlaArgAlaAlaVal 821
QY 997 CCGCGCTGAGCGCA-----GCGCTGAGGCGCGCACTGACGCGCAGCGCAGC--- 956
Db 822 ProAlaAlaThrProAlaGlyValAlaValAlaLeuGlyPro-----ArgGlyGlyAla 839
QY 955 -----TGGCGCGCGCGG----- 944
Db 840 AlaGlyLysLeuCysTYrAlaIleProGlyHisLeuValCysProAlaGlyAspArgAlaAla 859
QY 943 -----CGTGGG----- 938
Db 860 IleProCysPheGlnArgGlyProGlyAlaProGlyProGlyTYrGlyProGlySer 879
QY 937 AGATGACATGCGCTGTG----- 920
Db 880 GlyGlyGlyCysHisLeuProGluGlyAlaGlyTYrLeuThrAlaAlaGlyAlaGlyGlu 899
QY 919 -----GCCCGGCTTCCTGCGCGCGTGCAGACCGCGCAGAGATCTGGGCGCTCAGGA 869
Db 900 TrpProProSerGlyProAlaAlaThrAlaLeuLeuProAlaGlyTYrMetGlyGlyGly 919
QY 868 -----GGGCGACGTTGACT-----CAG 851
Db 920 LeuCysSerAlaGlyArgSerTYrAlaGlySerGlyGlyCysAlaGlyCysThrGlyPro 939
QY 850 GCACCTAGCGCAGCT-----CGTAGCTCCGTGTCGCGGT 818
Db 940 AlaAlaGlyProArgAlaGlnCysArgHisIleLeuProGlyAspAlaGlyProGlyProGly 959

```
QY      817 CGAGCGCGCCGATCCAGTCCG----- 794
      |||::|||
Db      960 ProGlyGlnPro----SerSerProAlaArgMetGlyProLeuProGlyProLeuProArg 978
QY      794 ----- 794
Db      979 AlaArgGluGlnAspProAlaThrProGlyArgGlyGlyGlnProThrGlyLeuProThr 998
QY      793 -----TGGCGTTCC-----CTGGCA 779
      |||::|||
Db      999 ThrAlaGlyArgArgCysGlnGlnIleProValGlyAlaProGlnProLeuAla 1018
QY      778 GCGTGGCGGTCTTGAGAGCCCGCGCTGGCGCA-----CGAGCTCCAGCA 728
      |||::|||
Db      1019 GlnProGlnLeuLeuAlaAlaProGlnGlnProTyrAlaThrAlaSerProIleProLeu 1038
QY      727 CATAGTACCGCGAGTCTGGCGGTGAGAGGGGTGGCCAGGCGGAGCGGAGCGCTGGAG 668
      |||::|||
Db      1039 LeuProGlyProMetTyrArgGlyLeuGly----- 1048
QY      667 TGATCTCCGTGGATGAGCTGCGCGCGCATCCGAGMATGAGCCCGCTGAGCTTT 608
      |||::|||
Db      1049 -----ArgGlyProAlaGlySerArgSerArgGlyGlnAlaProLysSer--- 1063
QY      607 GGAATATGATGTCCAGGTATCCAGTCCAGAAAGTGCAGGTCTTCTCGGCGAGGGCTG 548
      |||::|||
Db      1064 -----CysAla---AspProArgPro-----GlyGlyHisGlnHisGlyGly 1076
QY      547 AGGCGAGCGGTGACAGCTCCAGGAAGTGGCTGGCGCGGTGCGAGCAATGAACAGGTGA 488
      |||::|||
Db      1077 ArgGlnAspValLeuThrThrGly-----ThrArg--- 1086
QY      487 CGCCGAGCTCTTGAAGCTCTGATGAGGGAGC----- 455
      |||::|||
Db      1087 AlaAlaGlyPro-----GlyGlyAlaArgArgThrLeuGlySerArgArg 1101
QY      454 ---CCAGAGGTGCTGGAGCCGCGCATCTGTACCCAGCACAGCACTTTGGGCAACC--- 401
      |||::|||
Db      1102 ProProAspGlyAlaGlnAlaGlyHisGlnCysProAlaAlaAlaGlyAlaAlaAspCys 1121
QY      400 -----CTGGCGCGGACCGTGAATGCTTACAGCAACAGCTTCTTGGCATTAAGCA 350
      |||::|||
Db      1122 LeuThrArgLeuArgGlyHis-----LeuGlnAlaSerAlaThrProTyrAlaAlaAsp 1139
QY      349 GCG---CCAGGCGCAGTGGGTGTACCCA---TGGCTGGCGAGAGCAGCACCGCAT 296
      |||::|||
Db      1140 AlaThrSerGlyHisIleuGlyCysCysProGluCysProGly---LysAlaSerGlnLeu 1158
QY      295 CTTGGGAGCTTACCCGAGCTGTGCTGGCGAGAGGGGAACTCGGTGTATGGCCGACTGC 236
      |||::|||
Db      1159 Pro-----ProAspThrLeuSerAlaGlyAlaSerGlyLeuArg-----His 1172
QY      235 CCAGCTGCACAGACTGGCAGCGAGGAGCCCGGCTGCCAGGGGCACTGAGCCA----- 182
      |||::|||
Db      1173 ProProProThrHisTyrGlyLeuLeuProSerProArgGlyProValGlnProLeuCys 1192
QY      181 -----CCAGCTGCCCAACAATCCCGAACCAGGAGACTGTAGTAGAGACGC 131
      |||::|||
Db      1193 ThrValArgGlnAlaProThrGlnThrGlyGlu----- 1203
QY      130 TGGCTAGCTGTCCAGCAGGA---ACATCAGTCCCTCGGGGGCTGATGTCTGTGAC 74
      |||::|||
Db      1204 TrpSerGlyCysAlaGlnSerLeuLysGlnGlyLeuHisGlyTyrTrpProLeu----- 1221
QY      73 CGGCTCCGGCGCGCTCCGCGCAGAGCGCCGCGGAGAGCTCAGAGG----- 29
      |||::|||
Db      1222 -----ProAlaProSerProAlaAlaAlaSerGlyThrAlaAspSerVal 1236
QY      28 -----CCAGCGCAGGCGCG 14
      |||::|||
Db      1237 TrpAlaAlaProGlyGlyAlaPro 1244
```

```
US-10-131-826A-294
; Sequence 294, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RJC128
; CURRENT APPLICATION NUMBER: US/10/131, 826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See file wrapper or PAM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 294
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-294

Alignment Scores:
Pred. No.: 4.11e-08 Length: 915
Score: 253.50 Matches: 62
Percent Similarity: 50.0% Conservative: 36
Best Local Similarity: 31.6% Mismatches: 85
Query Match: 10.7% Indels: 13
DB: 6 Gaps: 3

US-10-699-035A-5 (1-1254) x US-10-131-826A-294 (1-915)
QY      46 GGGCTGGCGGAGCGCGCGGAGCGCGGTCCAGCAGATCAGCCCCCGAGGGGACCTG 105
      |||::|||
Db      45 AlaLeuLeuGlnSerSerCysGluAsn-----LysArgAlaAspLeu 58
QY      106 ATGTTCTGCTGAGACAGCTACAGCGGTCTCTCACTACAGATTCTCCCGGGTTCGGAG 165
      |||::|||
Db      59 ValPheIleIleAspSerSerArgSerValAsnThrHisAspTyrAlaLysValLysGlu 78
QY      166 TTGTGGGCGAGCTGTGAGTCACTGAGCCCTGGGGACCGGGGCGCTGCTGCGCACTG 225
```

```

Db      79 PheIIeValAspIIeLeuGlnPheLeuAspIIeGlyProAspValThrArgValGlyLeu 98
      226 GGCACGTCGCGAGTCGCGCCATACACCGAGTCCCTTCGCGCAGACAGCTCGGCTGAG 285
      99 LeuGlnTYRGISeThrValIysAsnGluPheSerLeuIysThrPheIysArgIysSer 118
      286 GCTGCCAGAGATCGCGTGCCTCTCTGCCAGCGATGGGTGACACCCACACTGGCCTG 345
      119 GluValAGluArgAlaValIysArgMetArgHisLeuSerThrGlyThrMetThrGlyLeu 138
      346 GCGCTGCTATGACCAAGAACAGCTGTTGCTGAAGCATCAGTCCCGCGGCA----- 399
      139 AlaIleGlnTYRAlaLeuAsnIleAlaPheSerGluAlaGluGlyAlaArgProLeuArg 158
      400 ---GGGGTCCCAAGTCTGGTGTGTCGAGACAGATGGCGGCTCCAGCAGCCTGTGGGC 456
      159 GluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSerValAla 178
      457 CCCCCCATGACAGAGCTCAGAGACCTGGGCGTCCACCGTTCATGTCAGACCGCGCA 516
      179 GluValAlaAlaIysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyValGlyGln 198
      517 GGCACCTTCCTGAGACTGTACCGCTGCTCAGCCCTCGCGGAGAGACCTG----- 570
      199 ValAspPheAsnThrLeuIysSerIleGlySerGluProHisGluAspHisValPheLeu 218
      571 -----CACTTGTGACGTGATGACATCATGTCACAGAG 612
      219 ValAlaAsnPheSerGlnIleGlnThrLeuThrSerValPheGlnIys 234

```

RESULT 6
US-11-113-424-39
; Sequence 39, Application US/11113424
; Publication No. US20050260713A1

```

; GENERAL INFORMATION:
; APPLICANT: Gangoli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; PRIOR FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 39
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-39

```

Alignment Scores:
Pred. No.: 4,09e-08 Length: 956
Score: 253.50 Matches: 62
Percent Similarity: 50.0% Conservative: 36
Best Local Similarity: 31.6% Mismatches: 85
Query Match: 10.7% Indels: 13

```

DB:      7 Gaps:      3
US-10-699-035A-5 (1-1254) x US-11-113-424-39 (1-956)
      46 GCGCTGGGCGGAGGCGGCGCGGCTCCACAGATACAGCCCGGAGGAGCCTG 105
      45 AlaLeuLeuGluSerSerCysGlnAsn-----LysArgAlaAspLeu 58
      106 ATGTTCTCTGTCGACAGCTCAGCCAGCGCTCTCTCACTACAGATTCTCCCGGTTGAGAG 165
      59 ValPheIleIleAspSerSerArgSerValAsnThrHisAspTYRAlaIysValIysGln 78
      166 TTTCGGGGCAGCTGTGTGCTTCACCTGCCCTCGGACCGGGGCGCTCGTCCAGTCTG 225
      79 PheIIeValAspIIeLeuGlnPheLeuAspIIeGlyProAspValThrArgValGlyLeu 98
      226 GGCACGTCGCGAGTCGCGCCATACACCGAGTCCCTTCGCGCAGACAGCTCGGCTGAG 285
      99 LeuGlnTYRGISeThrValIysAsnGluPheSerLeuIysThrPheIysArgIysSer 118
      286 GCTGCCAGAGATCGCGTGCCTCTCTGCCAGCGATGGGTGACACCCACACTGGCCTG 345
      119 GluValAGluArgAlaValIysArgMetArgHisLeuSerThrGlyThrMetThrGlyLeu 138
      346 GCGCTGCTATGACCAAGAACAGCTGTTGCTGAAGCATCAGTCCCGCGGCA----- 399
      139 AlaIleGlnTYRAlaLeuAsnIleAlaPheSerGluAlaGluGlyAlaArgProLeuArg 158
      400 ---GGGGTCCCAAGTCTGGTGTGTCGAGACAGATGGCGGCTCCAGCAGCCTGTGGGC 456
      159 GluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSerValAla 178
      457 CCCCCCATGACAGAGCTCAGAGACCTGGGCGTCCACCGTTCATGTCAGACCGCGCA 516
      179 GluValAlaAlaIysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyValGlyGln 198
      517 GGCACCTTCCTGAGACTGTACCGCTGCTCAGCCCTCGCGGAGAGACCTG----- 570
      199 ValAspPheAsnThrLeuIysSerIleGlySerGluProHisGluAspHisValPheLeu 218
      571 -----CACTTGTGACGTGATGACATCATGTCACAGAG 612
      219 ValAlaAsnPheSerGlnIleGlnThrLeuThrSerValPheGlnIys 234

```

RESULT 7
US-10-821-234-1431
; Sequence 1431, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: PC_SEQ_genes Version 1.0
; SEQ ID NO 1431
; LENGTH: 1366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1431

Alignment Scores:
Pred. No.: 2.5e-06 Length: 1366
Score: 221.50 Matches: 135
Percent Similarity: 34.8% Conservative: 28
Best Local Similarity: 28.8% Mismatches: 173
Query Match: 9.3% Indels: 133


```

OY 93 -----CCGAGGGACCTGATGTTCTCTGAGACGCTC 125
DB 612 LysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArg 631
OY 126 AGCCAGCGCTCTCACTACGAGTTCCCGGGGTGGAGTTTGGGGGACCTGCTGGC 185
DB 632 GlyGluGln-----GlyProAlaGlySerProGlyPhe 642
OY 186 TCCACTGCCCCCTGGGACCCGGGGCCCTGCGTCAGTGTGCACTGTCAGTGTGGCC 245
DB 643 GlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProGlyGlnGlnGly 662
OY 246 ATACACCAAGTTCCCTTGGCCAGACAGCTGGGTGAGCTGCCACAGATGCGGTGGC 305
DB 663 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 674
OY 306 TGCTTCTGCCAGCGCATGGGTGACACCCACACTGGCTGGGCTGTGTATGCAAGGA 365
DB 675 -----AlaArgGly--GluArgGlyPheProGlyGluArgGlyValGlnGly 689
OY 366 ACAGCTGTTTGTGAAGCATCAGTGTCCCGGCC-----AGGGGTGCCAAGTGTGGT 419
DB 690 ProPro-----GlyProAlaGlyProArgGlyAlaAsnGlyAlaPro 703
OY 420 GTGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCCCCCAGTCA--GAGCTCA 476
DB 704 GlyAsnAsp-----GlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGln 721
OY 477 GGA-----CCTGGGCGCTACCGTGTCTATTGTACGACACCGGCCAGAGAACTTCTGAGCT 533
DB 722 GlyAlaProGlyLysGlnGlyMetProGlyGluArgGlyAlaAlaGlyLysProGlyPro 741
OY 534 GTCAACCGCTGCTCAGCCCTGCGGAGAAGCACTGTCATTGTGACGTGATGACCT 593
DB 742 -----LysGlyAspArg 745
OY 594 GCACATCATTTTCACAGA--GCTGAGGGCTTCATTCTCCGAGTCCGCC----- 641
DB 746 GlyAspAlaGlyProLysGlyLysAspGlySerProGlyLysAspGlyValAlaArgGlyLeu 765
OY 642 -----GCAGACGCTCCATGCGACGGAATCACGTCGAG 674
DB 766 ThrGlyProIleGlyProProGlyProAlaGlyAlaPro-----GlyAspLysGlyGlu 783
OY 675 CGGCTTCCGCTGCTGCGCACCCCTGCTGACCGGACGCTGGGCTACTATGTGCTGA 734
DB 784 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 797
OY 735 GCTGGTGGCCAGCGCCAGCGGGGCTGCAGACCGCAGAGCTGCCAGGAAAGCCAC 794
DB 798 AlaProGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGlyProPro 817
OY 795 GGACTGGATCTGGGCGCGCT-----CGACCGGACACGAGATTACAGCTGGC 842
DB 818 GlyAlaAsp--GlyGlnProGlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGly 836
OY 843 GCTAGTCTGATCCAGCTGCGCTCTGAGGCCCGCCAGACTCTGCG----- 890
DB 837 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 851
OY 891 -----GTCGCGACCGCGCCAGAGAGAGCGCG----- 917
DB 852 ProIleGlyAsnValGlyAlaProGlyAlaLysGlyAlaArgGlySerAlaGlyProPro 871
OY 918 -----GCCAGAGCGCATCGTCACTCTCCAGCGCGCGC----- 950
DB 872 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 891
OY 951 GCGAGCGCTCCGCGTGAAGTTGGGCCCGACGCTGGGCTGAGCGCGCGCTGGCTACGA 1010
DB 892 AlaGlyProProGlyProProGlyProAlaGly-----LysGlnGlyGlyLysGlyPro 909
OY 1011 CGTGACGTTCCGGCGCGCTGCGGGGCGGGAGCGGACGCGGCT----- 1052

```

```

DB 910 ArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyProProGly 929
OY 1053 -----GAGGTGCTCCGCGGCGCGCAACTGACACCGCTGAGG 1091
DB 930 ProAlaGlyGluLysGlySerProGlyAlaAspGlyPro-----AlaGlyAla 945
OY 1092 CCTGGC-----CCGAGCACCGC-----CTACCTGCT 1118
DB 946 ProGlyThrProGlyProGlnGlyLeAlaGlyGlnArgGlyValAlaGlyLeuProGly 965
OY 1119 GACCTGACCGCGC-----CTTCGCTCGGCGCGCCAGAGACCGGCTGCC 1166
DB 966 GlnArgGlyGluArgGlyPheProGlyLeuPro--GlyProSerGlyGluProGlyLys 984
OY 1167 CAAGCCTGACAGCGCCGACGCGCGCGCCAGCGCGCGCGCGCGCGCGAC 1226
DB 985 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro--ProGlyP 999
OY 1227 CCGGCGGACCGCGACCGCTGAGC 1249
DB 999 romerGlyProProGlyLeuAla 1006

RESULT 11
US-11-186-284-28
; Sequence 28, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-28

Alignment Scores:
Pred. No.: 8,52e-06 Length: 1464
Score: 212.00 Matches: 140
Percent Similarity: 33.6% Conservative: 31
Best Local Similarity: 27.5% Mismatches: 168
Query Match: 8.9% Indels: 170
DB: 7 Gaps: 33

US-10-699-035A-5 (1-1254) x US-11-186-284-28 (1-1464)
OY 15 GCGGCTCGG-----CCTGCGCTGAGCTTGGCGGTGGC--- 47
DB 572 GlyAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGlu 591
OY 48 ---GCTGCGCGGAGCGCGCGCGGAGCGGCTCACCC-----AGCATAGCCCC----- 92

```

```

Db      522 ProGlyLyAspAlaGlyGluArgGlyValProGlyProProGlyAlaValGlyProAlaGly 611
Qy      93 -----CCGAGGGGACCTGATGTTCTCTCTGAGACGTC 125
Db      612 LysAspGlyGluAlaGlyAlaGlyAlaGlyProProGlyProAlaGlyProAlaGlyGluArg 631
Qy      126 AGCCAGCGCTCTCTCACTACGAGTTCTCCGGGTTCCGGAGTTGTGGGCGACGCTGTGCG 185
Db      632 GlyGluGln-----GlyProAlaGlySerProGlyPhe 642
Qy      186 TCCACTGCCCCCTGGGCGACCGGGCCCTGCGTCCCACTCTGTGCGACAGTGTGCGGCC 245
Db      643 GlnGlyLeuProGlyProAlaGlyProGlyGluAlaGlyLysPheProGlyGlnGlnGly 662
Qy      246 ATACA CCGAGTTCCCTTCCGCGACAGACGCTGCGGTGAGCGCGCCGACGATGCGGTGCG 305
Db      663 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 674
Qy      306 TGCTTCTGCGCCAGCGCATGAGTGAACCCACACTGCGCTGCGCGCTGTATGCCAAGA 365
Db      675 -----AlaArgGly-----GluArgGlyPheProGlyGluArgGlyValGlnGly 689
Qy      366 ACAGCTGTTGCTGAAGCATCAGGTCGCCGCC-----AGGGGTGCCCAAGTGTGCT 419
Db      690 ProPro-----GlyProAlaGlyProArgGlyAlaAsnGlyAlaPro 703
Qy      420 GTGGGAGACAGATGCGCGCTCCAGCGACCCCTGTGGGCCCCCCATCA---GGAGCTCAA 476
Db      704 GlyAsnAsp-----GlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGln 721
Qy      477 GGA---CCTGGGCGCTCACCGGTGTTATGTCAGACCGGCGCAGGCACTTCTGTGAGCT 533
Db      722 GlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyPro 741
Qy      534 GTACGCGCTGCTCAGCCCCCTGCGGAGAACCTGCACTTTGTGAGACTGATACCT 593
Db      742 -----LysGlyAspArg 745
Qy      594 GCACATCATTTGCCAAGA---GCTGAGGGGCTCCATTTCGGAGTGGCC----- 641
Db      746 GlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGlyValArgGlyLeu 765
Qy      642 -----GCAGCAGCTCATGTCACCGAGATCAGCTCCAG 674
Db      766 ThrGlyProLysGlyProProGlyProAlaGlyAlaPro-----GlyAspLysGlyGlu 783
Qy      675 CGGCTTCCGCTGGCTGCGCACCCCTGCTGACCGGACCTGCGGCTACTATGTGCTGA 734
Db      784 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 797
Qy      735 GGTGTGCCACGCGCCGCGGCGGCTGCCAAGACCGCAGCTGCCAGGAAGCCAC 794
Db      798 AlaProGlyAspAlaGlyGluProGlyProProGlyProAlaGlyPheAlaGlyProPro 817
Qy      795 GAGACTGATCTCGGCGCGGCT-----CGACCCGAGCACCGACTACGACGCTGCG 842
Db      818 GlyAlaAsp---GlyGlnProGlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGly 836
Qy      843 GCTAGTGCCTGAGTCCACGTCGCTCTCTGAGGCCCCAGATCTGCG----- 890
Db      837 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 851
Qy      891 -----GGTGGGACGCGGCGGAGGAGGAGCGCG----- 917
Db      852 ProLysGlyAsnValGlyAlaProGlyAlaLysGlyAlaArgGlySerAlaGlyProPro 871
Qy      918 -----GCCAGAGCGCATGTCATCTCCACGCGCGCGC----- 950
Db      872 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 891
Qy      951 GCGCAGCTTCCGCGTGAATTGGGCCCCAGCGCTGGGCTAGCCGCGCGCTCGGCTAACCA 1010

```

```

Db      892 AlaGlyProProGlyProProGlyProAlaGly-----LysGluGlyGlyLysGlyPro 909
Qy      1011 CDTGCAATTCCGGCCGCTCCGCGGCGGAGCGACGCGGT----- 1052
Db      910 ArgGlyGluThrGlyProAlaGlyLysProGlyGluValGlyProProGlyProProGly 929
Qy      1053 -----GAGGTGCCCCCGGCGCAACTGACCACTGACGAG 1091
Db      930 ProAlaGlyGluLysGlySerProGlyAlaAspGlyPro-----AlaGlyAla 945
Qy      1092 CCTGCG-----GCCGGGCA CCGC-----CTACCTGCT 1118
Db      946 ProGlyThrProGlyProGlnGlyAlaGlyGlnArgGlyValValGlyLeuProGly 965
Qy      1119 GACCGTGA CCGCGC-----CTTCCGCTCGGCGCGCGAGCGCGCTGCTCCG 1166
Db      966 GlnArgGlyGluArgGlyPheProGlyLeuPro---GlyProSerGlyGluProGlyLys 984
Qy      1167 CAAGCTGTCACGCGCCGCGCGCGCGCGCCACGCGCCGCTGCTCCGCGCGCGC 1226
Db      985 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro--ProGlyP 999
Qy      1227 CCGGAGACCGCGCAGCCGCGAGC 1249
Db      999 romecGlyProProGlyLeuAla 1006

RESULT 12
US-11-021-603-2
; Sequence 2, Application us/11021603
; Publication No. US20060003954A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/11/021, 603
; CURRENT FILING DATE: 2004-12-21
; PRIOR APPLICATION NUMBER: US/10/149, 352
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRF
; ORGANISM: Homo sapiens
US-11-021-603-2

Alignment Scores:
Pred. No.: 8,52e-06 Length: 1464
Score: 212.00 Matches: 140
Percent Similarity: 33.6% Conservative: 31
Best Local Similarity: 27.5% Mismatches: 168
Query Match: 8.9% Indels: 170
DB: 7 Gaps: 33

US-10-699-035A-5 (1-1254) x US-11-021-603-2 (1-1464)
Qy      15 GCGCGCTCGG-----CCTGGCCCTGAGCTTCCGGCTGCG--- 47
Db      572 GlyAlaArgGlyGluAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGlu 591
Qy      48 ---GCTGGCGGAGCGCGGCGGAGCGCGGTCCACC-----AGCATGAGCCCC----- 92
Db      592 ProGlyLysAspGlyGluArgGlyValProGlyProProGlyAlaValGlyProAlaGly 611
Qy      93 -----CCGAGGGGACCTGATGTTCTCTCTGAGACGTC 125
Db      612 LysAspGlyGluAlaGlyAlaGlyAlaGlyProProGlyProAlaGlyProAlaGlyGluArg 631
Qy      126 AGCCAGCGTCTCTCACTACGAGTTCTCCGGGTTCCGGAGTTGTGGGCGACGCTGTGCG 185

```

```
Db      632 GlyGluGln-----GlyProAlaGlySerProGlyPhe 642
Oy      186 TCACATGCGCCGTCGACCGGGGCGCTGCGTCAGATGTCGACATGTCGAGCG 245
Db      643 GlnGlyLeuProGlyProAlaGlyProProGlyGlyAlaGlyLeuProGlyGlnGly 662
Oy      246 ATACACCGAGTTCCCTTCGCGCAGACAGCTCGGAGTGCCTCCAGAGATGCGGCG 305
Db      663 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 674
Oy      306 TCCTTCGCGCAGCGCATGCGTGAACCCACACTGCGCTGCGCTGCTATGCGCAAGA 365
Db      675 -----AlaArgGly-----GluArgGlyPheProGlyGluArgGlyValGlnGly 689
Oy      366 ACAGCTGTTGTCGAAGCATCAGATGCGCGCGC-----AGGGGTGCGCCAAAGTCTGCT 419
Db      690 ProPro-----GlyProAlaGlyProAlaGlyAlaAlaGlnGlyAlaPro 703
Oy      420 GTGGGTGACAGATGCGCGCTCCAGCGACCCCTGTGGGCCCCCGCATGCA--GAGCTCAA 476
Db      704 GlyAsnAsp-----GlyAlaGlySerGlyAspAlaGlyAlaProGlyAlaProGlySerGln 721
Oy      477 GGA---CCTGGGCGTCACCGTGTTCATTGTACAGACCGGCGCAGCAACTTCTGAGCT 533
Db      722 GlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyPro 741
Oy      534 GTACGCGCGTGCCTCAGCCCTGCGCAGAGACCTGCACTTTGTGAGATGATGACT 593
Db      742 -----LysGlyAspArg 745
Oy      594 GCACATCATTTGCCAAGA--GCTGAGGGGCTCCATTCCGCGATGCGGCGC----- 641
Db      746 GlyAspAlaGlyProGlySerGlyAlaAspGlySerProGlyAlaAspGlyValArgGlyLeu 765
Oy      642 -----GAGACAGCTTCATGCCACGGAATCAGCTCCAG 674
Db      766 ThrGlyProGlyProGlyProGlyProAlaGlyAlaPro-----GlyAspLysGlyGlu 783
Oy      675 CGGCTTCGCGCTGCGTCGACACCCCTGTCAGCCGACGCTGAGCTATGTCGTGGA 734
Db      784 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 797
Oy      735 GCTGTGCGCCAGCGCCGCGGCGCTGCAAGACCGCAGACGTCGCGAGAAAGCCAC 794
Db      798 AlaProGlyAspArgGlyGluProGlyProGlyProAlaGlyPheAlaGlyProPro 817
Oy      795 GGAAGTGAATCTGGCGCGGCT-----CGACCGGAGACAGACTACGACGTGGC 842
Db      818 GlyAlaAsp--GlyGlnProGlyAlaGlyGlyGluProGlyAlaAspAlaGlyAlaGlyGly 836
Oy      843 GCTAGTGCCTGAGTCCAGCGTCGCGCTCTGAGGCGCCCGACAGATCCTGCG-- 890
Db      837 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 851
Oy      891 -----GTCGCGCAGCGCGCCAGAGAGGCGCG-- 917
Db      852 ProGlyGlyAsnValGlyAlaProGlyAlaGlyAlaArgGlySerAlaGlyProPro 871
Oy      918 -----GCCAGAGCGCATCTGCTCCACGCGCGCGC----- 950
Db      872 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 891
Oy      951 GCGCAGCTCCGCGTGAATTTGGGCGCCAGGCGCTGAGCGCGCGCTCGGCTACCA 1010
Db      892 AlaGlyProProGlyProProGlyProAlaGly-----LysGlnGlyGlyGlyGlyPro 909
Oy      1011 CGTGCAATTGCGGCGCTCGCGGCGCGGAGCGCGCAGCGGT----- 1052
Db      910 ArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyProProGly 929
Oy      1053 -----GAGGTGCGCGCGCGCGCGCGCACTGCACCACTGCGAGGG 1091
```

```
Db      930 ProAlaGlyGlyLysGlySerProGlyAlaAspGlyPro-----AlaGlyAla 945
Oy      1092 CCTGCG-----GCCGGGACCGC-----CTACCTGCT 1118
Db      946 ProGlyThrProGlyProGlnGlyLeuAlaGlyGlnArgGlyValAlaGlyLeuProGly 965
Oy      1119 GACCGTGAACCGCGC-----CTTCGCTCGGGCGCGCGAGAGCGGCTGCTCCG 1166
Db      966 GlnArgGlyGluArgGlyPheProGlyLeuPro--GlyProSerGlyGluProGlyLys 984
Oy      1167 CAAGCGCTGACAGCGCCGCGCGCGCGCGCCAGCGCGCGCGCGCGCGCGCGCG 1226
Db      985 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro--ProGlyP 999
Oy      1227 CCGGGGACCGCGCAGCGCTGAGC 1249
Db      999 rometGlyProProGlyLeuAla 1006

RESULT 13
US-10-821-234-1096
; Sequence 1096, Application US/10821224
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIORITY FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pc_seq_genes Version 1.0
; SEQ ID NO 1096
; LENGTH: 1467
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1096

Alignment Scores:
Pred. No.: 8,52e-06 Length: 1467
Score: 212.00 Matches: 140
Percent Similarity: 33.6% Conservative: 31
Best Local Similarity: 27.5% Mismatches: 168
Query Match: 8.9% Indels: 170
DB: 6 Gaps: 33

US-10-699-035A-5 (1-1254) x US-10-821-234-1096 (1-1467)
Oy      15 GCGGCTCGG-----CCTGCGCTGAGCTTGCAGCTGAGC-- 47
Db      575 GlyAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProGlyValAlaAlaGlyGlu 594
Oy      48 ---GCTGCGCGGAGCGCGCGCGGCGGCTCAC--AGCATCAGCCCC----- 92
Db      595 ProGlyLysAlaGlyGluArgGlyValProGlyProProGlyValAlaGlyProAlaGly 614
Oy      93 -----CCGAGGGACCTGAGTCTCGCTGCGGACAGCTC 125
Db      615 LysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArg 634
Oy      126 AGCCAGCTCTCACTACAGAGTCTCCGCGGTTGCGGAGTTTGGGCGACAGCTGATGCG 185
Db      635 GlyGluGln-----GlyProAlaGlySerProGlyPhe 645
Oy      186 TCACATGCGCCCTGAGCAGCGGGGCGCTGCGTCAGATGTCGACATGTCGAGCGCGCC 245
Db      646 GlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProGlyGlnGly 665
Oy      246 ATACACCGAGTTCCCTTCGCGCAGACAGCTCGGAGTGCCTCCAGAGATGCGGCG 305
```

```

Db 666 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 677
Qy 306 TCCTTCGCCAGCCGATGGGTGACACCCAGACTGCGCGCTGTCTATGCCAAGA 365
Db 678 -----AlaArgGly-----GluArgGlyPheProGlyGluArgGlyValGlnGly 692
Qy 366 ACAAGCTGTTGCTGAAGCATCAAGTGTCCCGGCC-----AGGGGTGCCCAAGTGTGCT 419
Db 693 ProPro-----GlyProAlaGlyProArgGlyAlaAsnGlyAlaPro 706
Qy 420 GTGGGTGACAGATGGCGGCTCCAGCGACCCCTGTGGCCCCCATGCA---GGAGCTCA 476
Db 707 GlyAsnAsp-----GlyAlaGlySerGlyAlaProGlyAlaProGlySerGln 724
Qy 477 GGA---CCTGGGCGTCAACCGTGTCTATGTCAGACCGGCGGCACTTCCTGGAGCT 533
Db 725 GlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyPro 744
Qy 534 GTACGCGCTGCTCAGCCCTGCGGAGAACACTGCACTTTGTGAGCTGATCACT 593
Db 745 -----LysGlyAspArg 748
Qy 594 GCAACATCATTTGCCAAGA---GCTGAGGGGCTCCATTCTCGGATGCGGCC----- 641
Db 749 GlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGlyValArgGlyLeu 768
Qy 642 -----GCAGAGCTCATGTCACCGAGATCACTGCAG 674
Db 769 ThrGlyProLysGlyProProGlyProAlaGlyAlaPro-----GlyAspLysGlyLeu 786
Qy 675 CGGCTTCGCGCTGCGCTGCGCCACCTCTGACCGGCACTCGGGCTACTGTGCTGA 734
Db 787 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 800
Qy 735 GCTGTGCCAGAGCGCCAGCGGGGCTGCAGACCCGACGACTGCGAGAACCCAC 794
Db 801 AlaProGlyAspArgGlyLeuProGlyProGlyProAlaGlyPheAlaGlyProPro 820
Qy 795 GAAGTGTATCTGGCGCGCT-----CGACCGGACACGAGACTACGACGTGCG 842
Db 821 GlyAlaAsp---GlyGlnProGlyAlaGlyGlyLeuProGlyAspAlaGlyAlaGly 839
Qy 843 GCTAGTGCCTGAGTCCACGCTGCGCTCTGAGCGCCGACGATCTGCG----- 890
Db 840 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 854
Qy 891 -----GCTGCGACGCGCGCCAGAGGCGCG----- 917
Db 855 ProLysGlyAsnValGlyAlaProGlyAlaGlyAlaArgGlySerAlaGlyProPro 874
Qy 918 -----GCCAGAGCGCATGTCATCTCCGACGCGCGCC----- 950
Db 875 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 894
Qy 951 GCGCAGCCTTCGCGGTAGTTGGGCGCCAGCGCTGAGCTGACCGCGCGCTGACTACGA 1010
Db 895 AlaGlyProProGlyProProGlyProAlaGly-----LysGlnGlyGlyLysGlyPro 912
Qy 1011 CGTGCACTTCGGGCGCTGCGGGCGGAGGCGCAGCGGT----- 1052
Db 913 ArgGlyGlyThrGlyProAlaGlyArgProGlyGlyValGlyProProGlyProProGly 932
Qy 1053 -----GGAGGTGCGCGCGCGCGCACTGACACGCTGCAGG 1091
Db 933 ProAlaGlyGlyLysGlySerProGlyAlaAspGlyPro-----AlaGlyAla 948
Qy 1092 CCTGGC-----GCGGCGGACCGC-----CTACCTGCT 1118
Db 949 ProGlyThrProGlyProGlnGlyLeuAlaGlyGlnArgGlyValValGlyLeuProGly 968
Qy 1119 GACCGTGACCGCGC-----CTTCGCGCTGCGCGCGGAGAGCGCGCTGTCGCG 1166
Db 969 GlnArgGlyGlyLysGlyPheProGlyLeuPro---GlyProSerGlyGlyLeuProGlyLys 987

```

```

Qy 1167 CAAGCCTGCAAGCCCAAGCGCCGCGCCCGCCAGCCCGCTGCCCCCGCCAC 1226
Db 988 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro--ProGlyP 1002
Qy 1227 CCGGCGGACCGCGACCGCTGAGC 1249
Db 1002 romerGlyProProGlyLeuAla 1009

RESULT 14
US-11-000-463-243
; Sequence 243, Application US/11000463
; Publication No. US2005026423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radjic T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIPACN
; CURRENT APPLICATION NUMBER: US/11/000,463
; PRIOR FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-000-463-243

Alignment Scores:
Score: 9.71e-06 Length: 1464
Percent Similarity: 211.00 Matches: 128
Best Local Similarity: 31.2% Conservative: 30
Query Match: 8.9% Mismatches: 180
DB: 7 Gaps: 30

US-10-699-035A-5 (1-1254) x US-11-000-463-243 (1-1464)
Qy 1243 GCGTGGCGTCCCGGCGGTGCGGCGGACGCGGCGCGCGCGCGT 1184
Db 242 GlyGluArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 261
Qy 1183 CGGCGCTGCAAGCCTTGGCGG-----ACAGCGCGC 1154
Db 262 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGly 281
Qy 1153 TCTCGCGGCGGACGAGGCGGAGGCGGCGTCA-----CGGTCAACGAGTACGCGG 1106
Db 282 ProAlaGlyProLysGlyLeuProGlySerProGlyGluAsnGlyAlaProGlyGlnMet 301

```

```
QY 1105 TGCCCG-----GCCCGAGGCCCTGCA----- 1085
Db 302 GlyProArgGlyLeuProGlyGlyLuarGlyArgProGlyAlaProGlyProAlaGlyAla 321
QY 1084 -----GCGGTGAGATTGGGCGCGGGGACCTCCA-----CCGCGTGG 1043
Db 322 ArgGlyAsnAspGlyAlaThrGlyAlaAlaGlyProProGlyProThrGlyProAlaGly 341
QY 1042 CCTCCCGCGCGCGAGCGCGCAACTGACAGCTGTAGCGGAGCGCGCGGTGAGCCCA 983
Db 342 ProProGlyPheProGlyAlaValaGlyAlaGlyGlyAlaGlyProAln----- 358
QY 982 GCGCTGGGGCCCACTCAGCGGAGGCTCGCGCGCGCGGTGGAGATGACATGCGCT 923
Db 359 -----GlyProArgGlySerGlyGlyProGlyAlaGly----- 371
QY 922 CTGCGCCGCGGCT-----CCTCGGCGCGGTGGCACCGCGAGATCT---GGG 878
Db 372 -----GluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnProGly 389
QY 877 GCGTCAGGA-----GCGCAGCTTGGACTCAGGCGCACTAGCGCGCAGTGTGCTGCT 824
Db 390 AlaAspGlyGlnProGlyAlaGlyAlaGlyAlaGlyAlaGlyAlaGlyAla 408
QY 823 CCGGGT----- 818
Db 409 ProGlyPheProGlyAlaArgGlyProSerGlyProGlyGlyProGlyGlyProGly 428
QY 817 -----CGAGCGCGCGCGAGATCCAGTCCGTCGTCCTCCGACAGTCTGCT--- 773
Db 429 ProGlyGlyAsnSerGlyGlyProGlyAlaProGlySerGlyAlaAspThrGlyAlaGly 448
QY 772 -----GCGCTCTTGCAGCCCCG----- 755
Db 449 GlyGluProGlyProValGlyAlaGlyGlyProProGlyProAlaGlyGlyGlyGly 468
QY 754 ---GCTGGCGCTGGGACACGAGCTCAGCA-----CATACT 722
Db 469 ArgGlyAlaArgGlyGlyProGlyProThrGlyLeuProGlyProProGlyGlyLuarGly 488
QY 721 AGCCGAGTCTGGGCA-----GCGGGGTGGCGCGCGGCG 683
Db 489 GlyProGlySerArgGlyPheProGlyAlaAspGlyAlaGlyProGlyGlyProAla 508
QY 682 GGAAGCGGCTGGACGTATCTCGTGGACATGAGCTGCGCGCATCGCGAGATG 623
Db 509 GlyGluArg-----GlySerProGlyProAlaGlyProGlyGly 521
QY 622 AGCCCTCAGCTCTTGGACAATGATGTGAGGTATCCAGTCCACAAAGTGAAGTGTCT 563
Db 522 SerProGlyGlyAlaGly-----ArgProGlyGlyAlaGlyLeu 534
QY 562 TCTCGGAGGGGCTGAGGAGCGGCTGACAGCTCCAGAAAGTTCCTCGCGGTCTGA 503
Db 535 -----ProGlyAlaGlyLeuThrGlySerProGlySer----- 546
QY 502 CAATGAACGAGTGACGCCAGGTCTTGAATCTGTCATGG---GGGGGCCACAGGCT 446
Db 547 -----ProGlyProAspGlyLysThrGlyProProGlyProAlaGly 560
QY 445 CGCTGAGCGCGCATATGTACACCAACAGCATTTGGGACCCCTGGCGGACCTG 386
Db 561 GlnAspGlyArgProGlyProProGlyProGlyAlaArgGlyGlnAlaGlyValMet 580
QY 385 ATGCTTACAGCAACACTGTTCTTGGCAATAGACAGCGCGGAGTGGGTGTAC 326
Db 581 -----GlyPheProGlyProGlyGlyAlaAlaGlyGlyProGly----- 593
QY 335 CCATGCGCTGGGAGAGACGACCGCATCT----- 293
Db 594 ---LysAlaGlyGlyLuarGlyValProGlyProProGlyAlaValGlyProAlaGlyLys 612
QY 292 -----GGGAGCTCACCAGC---TGTCTGCGCGAAGGGA 257
```

```
Db 613 AspGlyLuarGlyAlaGlyGlyProProGlyProAlaGlyProAlaGlyGlyLuarGly 632
QY 256 ACTCGGTATATGCGGCACTGC-----CCAGTGCACACAC 221
Db 633 GlnGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProPro 652
QY 220 TGGCAGCAGAGCGCCCGGTGCCAGGAGCACTGAGCCACCAACTCC 161
Db 653 GlyLuarGlyLys-----ProGlyGlyGlyValProGlyAspLeuGlyAlaPro 670
QY 160 GAACCCGGGAGAACTCGTAGAGAGA-----CGGTGCTGACGTCTCA 116
Db 671 GlyProSerGlyAlaArgGlyGlyLuarGlyPheProGlyGlyLuarGlyValGlnGlyPro 690
QY 115 GCAGGAATCAGATGCTCCCTCGGGGGGCTGATGTGTGACCGCGCTCCGCGCTCC 56
Db 691 ProGlyProAlaGlyProArgGly-----AlaAsnGlyAlaProGlyAsnAsp 706
QY 55 GCGCCAGCGCCAGCGCAAGCTCAGGCGCGAGCGCGCTCCAGGAGCA 2
Db 707 Gly---AlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGlnGlyAla 723

RESULT 15
US-11-186-284-28
/ Sequence 28, Application US/11186284
/ Publication No. US2005026493A1
/ GENERAL INFORMATION:
/ APPLICANT: Millennium Pharmaceuticals, Inc.
/ APPLICANT: Berger, Allison
/ APPLICANT: Guillemette, Tracy L.
/ APPLICANT: Kamatkar, Shubhangi
/ APPLICANT: Schlegel, Robert
/ APPLICANT: Monahan, John E.
/ APPLICANT: Thibodeau, Stephen N.
/ APPLICANT: Burgart, Lawrence J.
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
/ TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
/ FILE REFERENCE: MPM01-029P2RNM
/ CURRENT APPLICATION NUMBER: US/11/186,284
/ CURRENT FILING DATE: 2005-07-21
/ PRIOR APPLICATION NUMBER: US/10/301,822
/ PRIOR FILING DATE: 2002-11-21
/ PRIOR APPLICATION NUMBER: US 60/339,971
/ PRIOR FILING DATE: 2001-12-10
/ PRIOR APPLICATION NUMBER: US 60/361,978
/ PRIOR FILING DATE: 2002-03-05
/ PRIOR APPLICATION NUMBER: US 60/381,988
/ NUMBER OF SEQ ID NOS: 228
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 28
/ LENGTH: 1464
/ TYPE: PRT
/ ORGANISM: Homo Sapiens
US-11-186-284-28

Alignment Scores:
Pred. No.: 9 71e-06 Length: 1464
Score: 211.00 Matches: 138
Percent Similarity: 31.2% Conservative: 30
Best Local Similarity: 25.7% Mismatches: 190
Query Match: 8.9% Indels: 180
DB: 7 Gaps: 30

US-10-699-035A-5 (1-1254) x US-11-186-284-28 (1-1464)
QY 1243 GCGTGGAGGTCCCGCGGCTCGGAGCGCGGCGGTGGGCGCGGCGCGGCGCT 1184
Db 242 GlyLuarGlyGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 261
QY 1183 CGGCGGTGACGCGCTTGGCGG-----ACAGCGCGC 1154
```

```

Db      262 ProGlyMetCysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGly 281
QY      1153 TCTCGCGCGCCGAGAAAGCCGCGCGTCA-----CGATCACAGGTAGACGG 1106
Db      282 ProAlaGlyProLysGlyGluProGlySerProGlyGluAsnGlyAlaProGlyGlnMet 301
QY      1105 TGCCCG-----GCGCAGAGCCCTGCA----- 1085
Db      302 GlyProArgGlyLeuProGlyGluArgGlyArgProGlyAlaProGlyProAlaGlyAla 321
QY      1084 -----GCGTGTCAAGTTGCGCGCCGCGGCACTCA-----CCGCGTGG 1043
Db      322 ArgGlyAsnAspGlyAlaThrGlyAlaAlaGlyProProGlyProThrGlyProAlaGly 341
QY      1042 CCTCCCGCGCCGACGCGCCGACACTGCACTGATAGCCGAGCGCGCGGTAGACCCA 983
Db      342 ProProGlyPheProGlyAlaValaGlyAlaLysGlyGluAlaGlyProGln----- 358
QY      982 GCGCTGGGCGCCAACTCACGCGGAGCGCTGCGCGCGCGGTGGAGATGACATGCGCT 923
Db      359 -----GlyProArgGlySerGlnGlyProGlnGlyValaArgGly----- 371
QY      922 CTGCGCCGCGCT-----CCTCTGCGCGCGTGGCGCACCCGACGATCT---GGG 878
Db      372 -----GluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnProGly 389
QY      877 GCCTCAGGA-----GCGCAGCGTTGAGACTCAGGCACTAGCGCAGCGTGTAGTCCGTG 824
Db      390 AlaAspGlyGlnProGlyAlaLysGlyAlaAsnGly---AlaProGlyIleAlaGlyAla 408
QY      823 CCGGGT----- 818
Db      409 ProGlyPheProGlyAlaArgGlyProSerGlyProGlnGlyProGlyGlyProProGly 428
QY      817 -----CGAGCGCGGCCAGATCCAGTCCGTGGCGTTCCCTGGCAGCTGCT--- 773
Db      429 ProLysGlyAsnSerGlyGlnProGlyAlaProGlySerLysGlyAspThrGlyAlaLys 448
QY      772 -----GCGGTCTGACGCCCCG----- 755
Db      449 GlyGluProGlyProValaGlyAlaGlnGlyProProGlyProAlaGlyGlnGlyLys 468
QY      754 ---GCTGGCGCTGGCGCACCGACTCCAGCA-----CATAGT 722
Db      469 ArgGlyAlaArgGlyGlnProGlyProThrGlyLeuProGlyProProGlyGluArgGly 488
QY      721 AGCCCGAGTCTGGGTCA-----GCAGGGTGGCCAGGCCAGGC 683
Db      489 GlyProGlySerArgGlyPheProGlyAlaAspGlyValaAlaGlyProLysGlyProAla 508
QY      682 GGAAGCGGCTGAGCGATCTCGTGGCATGAGCTGCGCGCGCATCGCGAATGG 623
Db      509 GlyGluArg-----GlySerProGlyProAlaGlyProLysGly 521
QY      622 AGCCCTCAGCTCTTGACATATGATGTGACAGTCACTCAAGTCAAAAGTGACAGGTGCT 563
Db      522 SerProGlyGluAlaGly-----ArgProGlyGluAlaGlyLeu 534
QY      562 TCTCGCAGGGGCTGAGGACGCGCTGACAGCTCCAGAAATTGCTCGGCGGTCGTA 503
Db      535 -----ProGlyAlaLysGlyLeuThrGlySerProGlySer----- 546
QY      502 CATGAACA CGGTAGCGCCAGGTCTTAGCTCTGACATG---GGGGGCCACAGGGT 446
Db      547 -----ProGlyProAspGlyLysThrGlyProProGlyProAlaGly 560
QY      445 CGCTGAGCGCGCATCTGTACACCAACAGCACTTTGGGACCCCTGGCGGGCACTTG 386
Db      561 GlnAspGlyArgProGlyProProGlyProProGlyAlaArgGlyGlnAlaGlyValMet 580
QY      385 ATGCTTCAGCAACAGCTGTTCTTGACATAGACAGCCGCGCACGCTGGGTGTAC 326

```

```

Db      581 -----GlyPheProGlyProLysGlyAlaAlaGlyGluProGly----- 593
QY      325 CCATGCGCTGGCGAGAACGACCGACCGATCCT----- 293
Db      594 ---LysAlaGlyGluAlaGlyValProGlyProProGlyAlaValaGlyProAlaGlyLys 612
QY      292 -----GGCAGCTCACCCGAGC---TGTCGTGGCCGAAGGGA 257
Db      613 AspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArgGly 632
QY      256 ACTCGGTATGCGCGACTGC-----CCAGGTGACACGAGC 221
Db      633 GluGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProPro 652
QY      220 TGGACGACAGGCGCCCGGTGCCAGGGGCGATGAGCCACAGCTGCCACAAACTGCC 161
Db      653 GlyGluAlaGlyLys-----ProGlyGlnGlnGlyValProGlyAspLeuGlyAlaPro 670
QY      160 GAACCCGGGAGAACTGCTAGTAGAGA-----CGCTGGCTGAGCTGTCCA 116
Db      671 GlyProSerGlyAlaArgGlyGluArgGlyPheProGlyGluArgGlyValGlnGlyPro 690
QY      115 GAGGAACATCAGGTCCCTCGGGGGGCTGATGCTGTGTGACCGCGCTCGCGCTCC 56
Db      691 ProGlyProAlaGlyProArgGly-----AlaAsnGlyAlaProGlyAsnAsp 706
QY      55 GCGCCAGCGCCAGCGCAAGCTCAGGGCCAGGCGGACGCGCTCCAGGGAAGA 2
Db      707 Gly---AlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGlnGlyAla 723

```

Search completed: February 13, 2006, 13:54:37
 Job time : 43.2027 secs

THIS PAGE BLANK (USPTO)

[illegible]

F.352-433/Domain: fibronectin type III repeat homology <FN3>
 F.442-525/Domain: fibronectin type III repeat homology <FN3C>
 F.553-614/Domain: fibronectin type III repeat homology <FN3D>
 F.623-707/Domain: fibronectin type III repeat homology <FN3>
 F.741-823/Domain: fibronectin type III repeat homology <FN3>
 F.832-914/Domain: fibronectin type III repeat homology <FN3>
 F.922-1009/Domain: fibronectin type III repeat homology <FN3>
 F.1040-1205/Domain: von Willebrand factor type A repeat homology <vWA2>

Alignment Scores:
 Pred. No.: 1,15e-13 Length: 1888
 Score: 268.00 Matches: 64
 Percent Similarity: 50.5% Conservative: 34
 Best Local Similarity: 33.0% Mismatches: 80
 Query Match: 25.5% Indels: 16
 DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x S78476 (1-1888)

```

QY      4  GACCTGAGTGTCTCTGTCAGACGCTCAGCCAGCGTCTCTACTACGAGTTCTCCGGGTT 63
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      158 Aspllevalilleuvalaspllyserlyserpserilleglylaryphenasphleagleval 177
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      64  CGGAGTTTGTGGGGCAGCTGTGGCTCCACTGCCCCCTGGGCAACGGGGCCCTGGTCC 123
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      178 ArgleupheneulgluasneulvalserlAlapheasvalglysergluylstrargval 197
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      124 AGCTGTGTGACGTGGGGCAGTGGCCATACACCGAGTCCCCCTTGGCCACGACAGTCG 183
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      198 GlyleuAlagIntyserglyasproarglleglultrphisleuasmlatyrglythr 217
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      134 GGTAGAGCTGCCAGATGGCGGTGGTCTCTGCCACCGCATGGGTGACACCCACT 243
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      218 LysasplAlavalLeuasplAlavalArgasneuprotyrlyserglylansnthrLeuthr 237
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      244 GGCCTGGCGCTGCTCTATCCCAAGAACAGCTGTTTGTGAAGCATCAGGTGCCGCCA 303
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      238 GlyleuAlaleuthrlyrilleuengluasenserpheylrProgluAlaglyAlargPro 257
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      304 GGGGTGCCCAAGTGTGTGGTGGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCCCC 363
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      258 GlyValserlylleglyleuilethrAspGlyLysserglAspAspValillePro 277
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      364 CCCATGACGAGACTTAAAGACCTGGCGGTCCACCGTGTATTTGTGACGACCGGCGAGGC 423
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      278 ProAlalyAsnleuArgAspAlaglyllegluLeupheAlilleglylValysAsnAla 297
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      424 AACTCTCTGAGACTGTGACGCCGCTGCTCAGCCCTGCGCGAGAGACCTGTGCTGTG 483
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      298 AsplleasngluleuylsleuilealaserglurProaspserrhrhlsvalTYrAsnVal 317
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      484 --GACGTGAGTACCTGCACATCTGTCGA----- 513
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      318 AlaAspAphAsnphewetanserlileValgluglyLeuthrArgThrValCysSerArg 337
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY      514 -----GAGCTGAGGGGCTCATCTCGCG 537
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      338 ValgluglInglnlulysglullellysgllyThrleAla 351
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
  
```

RESULT 3
 AA0020
 collagen alpha 1(XII) chain precursor - chicken
 N:Alternate names: Fibrochimerin
 C:Species: Gallus gallus (chicken)
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 09-Jul-2004
 C:Accession: AA0020; A34485; B34485; A28037; S23814; S22554; S28811
 R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obayashi, K.
 A:Title: The complete primary structure of type XII collagen shows a chimeric molecule with
 a novel region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site
 A:Reference number: AA0020; MUID:92011862; PMID:1918137
 A:Molecule type: mRNA

A>Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' regions.

A:Reference number: A45974; MUID:93280195; PMID:8505337

A:Accession: A45974

A>Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-1747 <GER>

A:Cross-references: UNIPROT:P32018, UNIPARC:UPI0000173C47

A:Experimental source: embryo skin

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBIP:133365)

R:Apte, S.S.

submitted to the EMBL Data Library, March 1992

A:Reference number: S30085

A:Accession: S30085

A:Molecule type: mRNA

A:Residues: 1472-1660 <APT>

A:Cross-references: UNIPARC:UPI00000006A2; EMBL:X65122; NID:g62871, PIDN:CAA46238.1; PubMed:11525591

R:Trueb, J.; Trueb, B. 1992

Bur. J. Biochem. 207, 549-557, 1992

A>Title: Type XIV collagen is a variant of undulin.

A:Reference number: S22916; MUID:92339443; PMID:1339349

A:Accession: S22916

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>

A:Cross-references: UNIPARC:UPI0000173C48

R:Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsenmayer, T.F.; van der Rest, M.; Mayne, Bur. J. Biochem. 201, 333-338, 1991

A>Title: Cloning of a cDNA for a new member of the class of fibrin-associated collagens

A:Reference number: S17035; MUID:92037585; PMID:1935930

A:Accession: S17035

A:Molecule type: mRNA

A:Residues: 1472-1659 <GOR1>

A:Cross-references: UNIPARC:UPI0000173C49

A:Accession: S20833

A:Molecule type: protein

A:Residues: 1551-1570, 1593-1599, 1639-1667 <GOR2>

A:Cross-references: UNIPARC:UPI0000173C49

C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimeric

F:40-204/Domain: von Willebrand factor type A repeat homology <VWA1>

F:236-317/Domain: fibronectin type III repeat homology <FN3>

F:326-409/Domain: fibronectin type III repeat homology <FN3>

F:418-498/Domain: fibronectin type III repeat homology <FN3>

F:507-591/Domain: fibronectin type III repeat homology <FN3>

F:625-707/Domain: fibronectin type III repeat homology <FN3>

F:716-798/Domain: fibronectin type III repeat homology <FN3>

F:806-893/Domain: fibronectin type III repeat homology <FN3>

F:924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>

F:1111-1352/Domain: non-collagenous NC4 #status predicted <NC4>

F:1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>

F:1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Alignment Scores:

Pred. No.:	Score:	Length:
8.74e-13	257.00	1747
Percent Similarity:	50.0%	Matches: 62
Best Local Similarity:	32.0%	Conservative: 35
Query Match:	24.5%	Mismatches: 81
DB:		Indels: 16
		Gaps: 2

US-10-699-035A-1 (1-537) x A45974 (1-1747)

QY 4 GACCTAGTGTTCCTCGTGGAAGCTCAGCCAGGCTCTCACTACGAGATTCTCCGGGTT 63
||||| :|||:::||||| :|||::: |||

Db 42 AsplleValIleleValAspelYserTrpSerIIeglyArgPheasnPhelagLeuVal 61
||| |||::: |||::: |||::: |||

QY CGGAGATTTGTGGGAGCATGCTGGTGCCTCACACTGCCCTCGGAGCACGAGGCGCTTCGTGCC 123
||| |||::: |||::: |||::: |||

Db 62 ArgLeuheleuGlunshenValserArgPheasnValGlyserGluysrThrArgVal 81
||| |||::: |||::: |||::: |||

QY 124 AGTGTGTGCACGTGGGACAGTGCAGCCATAACCAGACTTCCTTCGAGCAGCACAGCTCG 183
||| |||::: |||::: |||::: |||

```

Db      82 G1yLeuAlaGlnTyrSerG1yAspProArgI1eolUtrpHisLeuAsnAlaTyrC1yThr 101
Qy      184 GGTGAGCGTCCCGAGATGGCGTGCCTTCTTCCCGACGCGATGGGTGAACCCACT 243
      :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      102 LysAspAlaValLeuAspAlaAspArgAsnLeuProTyrIySg1yG1yAsnThrLeuThr 121
Qy      244 GGCCTGGCGCGCTGGTCTATGCGCAAGAACACACTGTTTCTGTGAAGCATCGAGCCGGCA 303
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      122 G1yLeuAlaLeuThrTyrTr1leuG1uAsnCysPheIySProG1uAlaG1yAlaArgPro 141
Qy      304 GGGGTCCCAAAAGTGCCTGGTGGTGAGACAGATGGCGAGTCCAGCACTCTGGAGCCC 363
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      142 G1yValSerLysTr1leG1y1leu1leThrAspL1ySsSerG1nAspAspVal1lePro 161
Qy      364 CCCATCAGAGAGCTCAAGACACTGGCGCTCACCCGTGTTCAATGTCAGACCGCGGAGGC 423
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      424 AACCTTCCTGAGAGTGTACAGCGCGCTCAGCGCCCTCGCGAAGACACTGACATTGTG 483
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      182 AspTr1eAsnG1uLeuIySg1u1leAlaSerG1uProAspSerThrHisValIyAsnVal 201
Qy      484 ---GAGCTGAGTACCTGCACATCATGTTCCAA----- 513
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      202 AlaAspPheAsnPhetMetAsnSer1leValG1uCluIyLeuThrArgThrValCysSerArg 221
Qy      514 -----GAGCTGAGCGGCTCCATTCTCGCG 537
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      222 ValG1uG1uG1nG1uIySg1uTr1leG1yG1yTr1leAla1a1a 235

RESULT 5
A33809
cartilage matrix protein precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 09-Jul-2004
C:Accession: A33809; A26364
R:Kiss, I.; Deak, F.; Holloway Jr., R.G.; Delius, H.; Meubst, K.A.; Frimberger, E.; Argi
J. Biol. Chem. 264, 8126-8134, 1989
A:Title: Structure of the gene for cartilage matrix protein, a modular protein of the ex
grins, Von Willebrand factor, complement factors B and C2, and epidermal growth factor.
A:Reference number: A33809; PMID:89255246; PMID:2542265
A:Accession: A33809
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-493 <RIS>
A:Cross-references: UNIPROT:P05099; UNIPARC:UPI000004P1EF; GB:X12346; GB:X12347; GB:X1233
R:Argraives, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goetlinck, P.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 464-468, 1987
A:Title: Structural features of cartilage matrix protein deduced from cDNA.
A:Reference number: A26364; MUID:87092429; PMID:3025875
A:Accession: A26364
A:Molecule type: mRNA
A:Residues: 78-493 <ARG>
A:Cross-references: UNIPARC:UPI000004P1EF; GB:M14792; NID:9211545; PIDN:AAA8695.1; PID:
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
F:37-204/Domain: von Willebrand factor type A repeat homology <VWA1>
F:225-260/Domain: EGF homology <EGF>
F:270-434/Domain: von Willebrand factor type A repeat homology <VWA2>

Alignment Scores:
Pred. No.: 5.36e-12 Length: 493
Score: 247.50 Matches: 62
Percent Similarity: 51.5% Conservative: 22
Best Local Similarity: 38.0% Mismatches: 78
Query Match: 23.6% Indels: 1
Db: 2 Gaps: 1

US-10-699-035A-1 (1-537) x A33809 (1-493)
Qy      4 GACCTGATGTTCTGCTGAGACAGCTGACCGCGCTCTCACTACGAGTCTCCGGGTT 63
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      272 AspLeuValPheLeuIleAspG1ySerIySsSerValAlaArgProG1uAsnPhetG1uLeuVal 291

```

QY 64 CGGAGTTTGTGGGACGCTGCTCCACTGCCCCCTGGGACCGGGCCCTGCTGTC 123
 Db 222 LysylsPheilleAsnglnlleValGluSerLeuGluValIserGluIysGlnIalaGlnVal 311
 QY 124 AGTCGGGACAGTCGGGACGATACACGAGTTCCCTTGGCCAGACAGCTGCG 183
 Db 312 GilyleuValIglTryserserservalArgingluPheProleuGlyGlnhelsashn 331
 QY 184 GGTGAGGCTGCGGAGATGCGGTGCGTCTTCTGCGCCAGCGCATGGATGACCCACAT 243
 Db 332 LysylsAspIleleValIalaValIysylsMetalaTyrmegGluIyselyThrmethr 351
 QY 244 GCGCTGGCGCTGCTTATGACAGACAGCTTTGCTGAAACATCAAGTCCCGGCA 303
 Db 352 GlylnalaleuIysTytleuValIasperserPheiserllealaasnGlyIalaArgPro 371
 QY 304 GGGGCGCCCAAGTCGTGTGGGTGATGATGCGGCTCCAGGACCGCTTGAGGCCCC 363
 Db 372 GlyValProIysValGlyIleValIalPheThrAspGlyAsgerGlnAspTyrlleThrAsp 391
 QY 364 CCCATGACGAGCTCAAGACCTGGGCGCTGACCGTTCATTGTGACGACCGCGGAGGC 423
 Db 392 AlaIalaLysylsAlaLysAspIleuGlyPheArgMetPheAlaValGlyValGlyAsnIa 411
 QY 424 AACTTCCTGAGACTGTACAGCCGCTGCTGACGCCCTCGCCGAGACAGCTGACTTTGTG 483
 Db 412 ValGluAspGluLeuArgGluIleAlaIserGluProValAlaGluHis---TyrPheTy 430
 QY 484 GACGTGAT 492
 Db 431 ThrIalaAsp 433

RESULT 6

S66522
 cartilage matrix protein precursor - mouse

C:Species: Mus musculus (house mouse)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
 C:Accession: S66522

R:Ascoli, A.; Hauser, N.; Studer, D.; Paulsson, M.; Hirtl, L.; Bosze, Z.
 Eur. J. Biochem. 236, 970-977, 1996

A:Title: Cloning, sequencing and expression analysis of mouse cartilage matrix protein C
 A:Reference number: S66522; MUID:96270751; PMID:8665920

A:Accession: S66522

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-500 <AS2>

A:Cross-references: UNIPROT: P51942; UNIPARC: UP10000029480; EMBL: U55035; NID: g1163178; PI
 C:Gene: CPM

C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
 F:1-29/Domain: signal sequence #status predicted <SIG>

F:430-500/Product: cartilage matrix protein #status predicted <MAT>

F:231-266/Domain: von Willebrand factor type A repeat homology <VMA1>

F:227-262/Domain: EGF homology <EGF>

F:221-238,234-247,249-262/Disulfide bonds: #status predicted

Alignment Scores:

Pred. No.: 2,33e-11 Length: 500
 Score: 229.50 Matches: 57
 Percent Similarity: 52.48 Conservative: 29
 Best Local Similarity: 34.88 Mismatches: 75
 Query Match: 22.88 Indels: 3
 DB: Gaps: 1

US-10-699-035a-1 (1-537) x S66522 (1-500)

QY 4 GACCTGATGTTCTCTGACAGCTCAGCCAGCTCTCTCACTACGAGTTCTCCGGGTT 63
 Db 45 AspleuValAlaPheValAlaAspserValArgProValGluPheGluIysVal 64
 QY 64 CGGAGTTTGTGGGACGCTGCTCCACTGCCCCCTGGGACCGGGCCCTGCTGTC 123
 Db 65 LysValPheLeuSerGlnValIleGluSerLeuAspValGlyProAsnIalaThrArgVal 84

QY 124 AGTCGTGACAGTCGGGACGATGCGCATACCGAGTTCCCTTGGCCAGACAGCTGCG 183
 Db 85 GilyleuValIasnyrAlaSerThrValIysProGluPheProleuArgAlaHisGlySer 104
 QY 184 GGTGAGGCTGCGGAGATGCGGTGCGTCTTCTGCGCCAGCGCATGGATGACCCACAT 243
 Db 105 LysAlaSerLeuLeuGlnIalaValAlaArgIleGlnPheLeuSerThrGlyThrmethr 124
 QY 244 GCGCTGGCGCTGCTTATGACAGACAGCTTTGCTGAAACATCAAGTCCCGG--- 300
 Db 125 GlyleuAlalaLeuGlnPheAlaIleThrIysAlaLeuSerAspAlaGluGlyGlyArgAla 144
 QY 301 -----CGAGGCGTCCCAAGTCGTGTGGGTGATGATGCGGCTCCAGGACCT 354
 Db 145 ArgserProAspIleSerTyValIleValIalThrAspGlyArgProGlnAspser 164
 QY 355 GTGGGCCCCCAGACGAGACTCAAGACCTGGGCGCTGACCGTTCATTGTGACGACC 414
 Db 165 ValArgAspValIserGluArgAlaArgAlaIserGlyIleGluLeuPheAlaIleGlyLeu 184
 QY 415 GCGCGAGCACTTCTGAGACTGTGACCGCTGCTGACGCCCTCGCCGAGACAGCTG 474
 Db 185 GlyArgValAspIysAlaThrLeuArgGlnIleAlaIserGluProGlnAspGluHisVal 204
 QY 475 CACTTTGTGAC 486
 Db 205 AspTyValGlu 208

RESULT 7

A37979
 cartilage matrix protein precursor - human

C:Species: Homo sapiens (man)
 C:Date: 12-Jul-1991 #sequence_revision 12-Jul-1991 #text_change 09-Jul-2004
 C:Accession: A37979; B37979

R:Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; Sh
 J. Biol. Chem. 265, 19624-19631, 1990

A:Title: Structure and chromosomal location of the human gene encoding cartilage matrix
 A:Reference number: A37979; MUID:91060568; PMID:2246248

A:Accession: A37979

A:Molecule type: DNA

A:Residues: 1-496 <JEN>

A:Cross-references: UNIPROT: P21941; UNIPARC: UP1000004F1ED; GB: J05667

A:Accession: B37979

A:Molecule type: mRNA

A:Residues: 157-290, 'L', 292-496 <JEN>

A:Cross-references: UNIPARC: UP1000016A6E8; GB: M55683; GB: J05666; GB: J05667; NID: g180651;
 C:Gene: GDB: CRTM

A:Cross-references: GDB: 127280; OMIM: 115437

A:Map position: 1p35-1p35

A:Introns: 32/1; 147/3; 222/1; 264/1; 403/1; 454/1; 481/1

C:Complex: homotrimer

C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
 F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-496/Product: cartilage matrix protein #status predicted <MAT>

F:227-262/Domain: von Willebrand factor type A repeat homology <VMA1>

F:221-238,234-247,249-262/Disulfide bonds: #status predicted

Alignment Scores:

Pred. No.: 7.02e-11 Length: 496
 Score: 233.50 Matches: 62
 Percent Similarity: 50.38 Conservative: 27
 Best Local Similarity: 35.08 Mismatches: 87
 Query Match: 22.38 Indels: 1
 DB: Gaps: 1

US-10-699-035a-1 (1-537) x A37979 (1-496)

Qy	4	GACSTGATGTTCCGTGCTGGACGCTACACCGACGCTCTACCTACAGATTTCTCCGGGTT	63
Db	275	AspIeuValPheIeuIleAspGlySerIysSerValAlaProGluAsnIleVal	294
Qy	64	CGGAGATTTCGGGGACGCTGATGAGCTCCATGCGCCCTGGGACACGGGGGCTCTGGCC	123
Db	295	LysIysPheIleSerGlnIleValAspThrIleAspValSerAspIysLeuAlaGlnVal	314
Qy	124	AGCTGTGGACACGTGGGGACGTGGGCCATACACCGAGTTCCTTGGCCAGACAGCTCG	183
Db	315	GlyLeuValGlnIlyrSerSerSerValArgGlnGluPheProLeuGlyAlaGhrPheIthr	334
Qy	184	GGTAGAGCTGCCAGAGATCGGATCGGTGCTTCTGCCACGCGATGGGTGACACCGACAT	243
Db	335	LysIysAspIleIysAlaIleValAlaArgAsnMetSerTyrMetGlnIlyrGlyThrMetThr	354
Qy	244	GGCGCTGGGGCTGCTTATGCCAAGGAACAGCTTTGTGTGAAGATCAGAGTCCCGGCCA	303
Db	355	GlyAlaIleAlaLeuIlyrTyrLeuIleAspAsnSerPheThrValSerSerGlyAlaIlePro	374
Qy	304	GGGGTGCCTCAAGTGCCTGTGTGGGTGACACATGGCGGCTCCACGACCCCTGTGAGGCC	363
Db	375	GlyAlaGlnIlyrValGlyIleValPheThrAspGlyAlaGserGlnAspTyrIleAsnAsp	394
Qy	364	CCCATGCAGAGACTCAAGGACCTTGGGCGTCCCGTTCATTGTTCAGACCGCGGCCAGCC	423
Db	395	AlaIleAluIysValAluIysAspIleuGlyPheIlysmetPheAlaValGlyValGlyAsnAla	414
Qy	424	AACCTTCCTGGAGCTGTGACCGCGCTGCCACCCCT---GCCGAGAAGACCTTCACATTT	480
Db	415	ValGluAspGlnIleuAspGlnIleAlaSerGluProValAlaGlnIlyrPheTyrThr	434
Qy	481	GTGGACGTGCATGCATTCACATCATTTGTCCAAAGACGAGGGGCTCATTT	531
Db	435	AlaAspPheIysThrIleAsnGlnIleGlyIysIysLeuGlnIlyrIlySile	451

RESULT 8
A:54849
collagen alpha 1(VII) chain precursor - human
N:Alternate names: procollagen alpha 1(VII) chain
C:Species: Homo sapiens (man)
C:Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #text change 09-Jul-2004
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.
J. Biol. Chem. 269, 20256-20262, 1994
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain
A:Reference number: A54849; MUID:94327588; PMID:8051117
A:Accession: A54849
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2944 <CHR>
A:Cross-references: UNIPROT:002388; UNIPARC:UPI000017A138; GB:L02870; NID:G987124; PIDN:
R:Tanaka, T.; Takahashi, K.; Funakawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A:Title: Molecular cloning and characterization of type VII collagen cDNA.
A:Reference number: PH0844; MUID:92231902; PMID:1567409
A:Accession: PH0844
A:Molecule type: mRNA
A:Residues: 'ERR',340-475,'PALSTRASHSTLCWRATRWHPNCRGSHWTRACPCPNRPSAHRARAG',524-528,'C',
A:Cross-references: UNIPARC:UPI000017A139; DDBJ:D11152; DDBJ:D13694; NID:G453688; PIDN:H
A:Experimental source: Keratinocyte
A:Note: The authors translated the codon ACC for residues 394 and 397 as Tyr
R:Parrene, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A:Reference number: S16316; MUID:91334580; PMID:1871109
A:Accession: S16316
A:Molecule type: mRNA
A:Residues: 815-892,'E',894-1439 <PAR>
A:Cross-references: UNIPARC:UPI000016A722; GB:M65158; GB:S49017; NID:G180914; PIDN:AAA99
A:Experimental source: Keratinocyte
R:Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisyambh, P.S.; Cook, M.E.; Wright, J.
Invest. Dermatol. 99, 691-696, 1992

```

A:Title: Noncollagenous (NC1) domain of collagen VI resembles multidomain adhesion proc
A:Reference number: 156328; MUID:93107742; PMID:1146284
A:Accession: 156328
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'EPF', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>
A:Cross-references: UNIPARC:UPI000016B3AC; GB:SI1226; NID:g262308; P1DN:AA824637.1; P1D:
R:Selzler, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glatville, R.W.; Burgeson, R.E.
J. Biol. Chem. 264, 3822-3826, 1989
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase
A:Reference number: A30296; MUID:89139437; PMID:2537292
A:Accession: A30296
A:Molecule type: protein
A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041, '
A:Cross-references: UNIPARC:UPI000017A13A; UNIPARC:UPI000017A13B; UNIPARC:UPI000017A13C;
A:Note: two reported peptides cannot be reliably located
R:Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous i
A:Reference number: 148103; MUID:93271985; PMID:8459916
A:Accession: 184686
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 2395-2871, 'S', 2873-2944 <RE2>
A:Cross-references: UNIPARC:UPI000016A724; GB:L06862; NID:g386713; P1DN:AAA89196.1; P1D::
R:Christiano, A.M.; Ruynaenen, M.; Uitto, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser subs
A:Reference number: A55255; MUID:94224777; PMID:8170945
A:Contents: annotation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL7A1; EBRI; EBD1; EB
A:Cross-references: GDB:128750; OMIM:120120
A:Map position: 3p21.3-3p21.3
A:Note: defects in this gene can result in dominant and recessive dystrophic epidermolys
A:Note: there are 118 introns
C:Complex: type VII collagen is probably a homotrimer
C:Function:
A:Description: structural component of extracellular polymer associated with anchoring f
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
F:1-16/Domains: signal sequence #status predicted <SIG>
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F:17-1253/Domains: amino-terminal nonhelical #status predicted <NC1>
F:36-201/Domains: von Willebrand factor type A repeat homology <VWA1>
F:3231-318/Domains: fibronectin type III repeat homology <FN1>
F:327-413/Domains: fibronectin type III repeat homology <FN2>
F:414-502/Domains: fibronectin type III repeat homology <FN3>
F:508-593/Domains: fibronectin type III repeat homology <FN4>
F:598-683/Domains: fibronectin type III repeat homology <FN5>
F:686-771/Domains: fibronectin type III repeat homology <FN6>
F:776-862/Domains: fibronectin type III repeat homology <FN7>
F:864-952/Domains: fibronectin type III repeat homology <FN8>
F:954-1045/Domains: fibronectin type III repeat homology <FN9>
F:1052-1219/Domains: von Willebrand factor type A repeat homology <VWA2>
F:1170-1172/Region: cell attachment (R-G-D) motif
F:1189-1253/Region: cysteine/proline-rich
F:1254-2783/Region: interrupted helical
F:1334-1336/Region: cell attachment (R-G-D) motif
F:2008-2010/Region: cell attachment (R-G-D) motif
F:2553-2555/Region: cell attachment (R-G-D) motif
F:2784-2944/Domains: carboxyl-terminal nonhelical #status predicted <NC2>
F:3876-2929/Domains: animal Kunitz-type proteinase inhibitor homology <BPI>
F:3377-786,1109/binding site: carbohydrate (Asn) (covalent) #status predicted
F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F:2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental
F:2634,2802,2804/Disulfide bonds: interchain #status predicted
Alignment Scores: 1.92e-10 Length: 2944
Pred. No.: 227.50 Matches: 59
Score:

```


QY 475 CACCTTTGACGCGATGACCTGCATCATTTGCCAAGCGTGGGCTCCATT 531
Db 198 AsnLeuGluAanPheThrAlaLeuHISglYleValglAspLeuValAlaSerVal 216
RESULT 10
T46488
hypothetical protein DKFZp43J065.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C/Accession: T46488
R/Diesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A/Reference number: Z23035
A/Accession: T46488
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-741 <AAA>
A/Cross-references: UNIPROT:Q8NDE6; UNIPARC:UPI000016ACCB; EMBL:AL137638
A/Experimental source: adult testis; clone DKFZp43J065
C/Genetics:
A/Note: DKFZp43J065.1
Alignment Scores:
Pred. NO.: 2.97e-09 Length: 741
Score: 213.00 Matches: 58
Percent Similarity: 50.8% Conservative: 34
Best Local Similarity: 32.0% Mismatches: 83
Query Match: 20.3% Indels: 6
Gaps: 4
US-10-699-035a-1 (1-537) x T46488 (1-741)
QY 4 GACCTGATGTTCTGCTGGACAGCAGCTGCTCTCACTACGAGTTCTCCGGGTT 63
Db 440 AspLeuValPheValIleAspGlySerIleuGluValAanPheGluValVal 459
QY 64 CGGAGATTGTGGGCGACCTGTGCTCCACTGCCCTGGGCGCGGCGCTGCTCC 123
Db 460 LysGlnPheValIleThrGlyIleAspSerLeuThrIleSerProValAlaAlaVal 479
QY 124 AGCTGTCGACGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 183
Db 480 GlyLeuLeuGlnIYserThrGlnAlaHISThrGluPheThrLeuValAanPheSer 499
QY 184 GGTAGGCTGCCCGATGCGGTCGCTCTTCTGCGGCGGCGGCGGCGGCGGCGG 240
Db 500 AlaIysAspMetLysLysAlaVal---AlaHISMetLysIYrMetGlyLysGlySerMet 518
QY 241 ACTGAGCTGGCGCTGCTGTATGCCAAGACAGCTGTTGTGAAGATCAGTGGCCGG 300
Db 519 ThrGlyLeuAlaLeuLysHISMerPheGlnIYrSerPheThrGlnGlyAlaAlaArg 538
QY 301 CCA-----GGGTCGCCCAAGTCTGTGTGGTGGGCGGCGGCGGCGGCGGCGG 351
Db 539 ProLeuSerThrArgValProArgAlaAlaIleValPheThrAspGlyAlaAlaIleAsp 558
QY 352 CCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 411
Db 559 AspValSerGluTyrPalaSerLysAlaLysAlaIleValIleThrMetYrAlaValAla 578
QY 412 ACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 471
Db 579 ValGlyLysAlaIleGluGlnIYrMetGlnIYrMetGlnIYrMetGlnIYrMetGln 598
QY 472 CTGCACTTTGTG---GAGCTGATATCTGCATCATTTGCCAAGAGTGAAGGCGCTCC 528
Db 599 LeuPheYrAlaGluAanPheSerThrMetAspGlnIleSerGluLysLeuLysGly 618
QY 529 ATT 531
Db 619 Ile 619
RESULT 11

CGH03A
collagen alpha 3(VI) chain precursor [validated] - human
M/Contents: collagen alpha 3(VI) chain, splice form A9/N10(-)
C/Species: Homo sapiens (man)
C/Date: 21-Nov-1993 #sequence_revision 12-Nov-1999 #text_change 16-Aug-2004
C/Accession: A59140; S13679; S24465; A57083; S28776; S00245; C31952; C29848; S26510; S48
R/Chu, M.L.
submitted to GenBank, May 1998
A/Reference number: A59140
A/Accession: A59140
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-3176 <CHU>
A/Cross-references: UNIPROT:P12111; UNIPARC:UPI0000126DAF; GB:X52022; NID:G3127925; PIDN
R/Chu, M.L.; Zhang, R.Z.; Pan, T.; Stokes, D.; Conway, D.; Kuo, H.J.; Glanville, R.; May
EMBO J. 9, 385-393, 1990
A/Title: Mosaic structure of globular domains in the human type VI collagen alpha-3 chain
itors.
A/Reference number: S13679; MUID:90151612; PMID:1689238
A/Accession: S13679
A/Molecule type: mRNA
A/Residues: 1-30,237-313, 'CWM', 318-322, 'AR', 326-1815, 'FD', 1818-1819, 'ID', 1822-3176 <CH5>
A/Cross-references: UNIPARC:UPI0000173C14; EMBL:X52022; NID:G3127925
A/Accession: S24465
A/Molecule type: protein
A/Residues: 574-585; 965-973, 'X', 975-976; 1306-1325; 1361-1377; 1381-1401; 1473-1506, 'X', 1508
-1962, 'X', 1964-1965; 2018-2037; 2374-2410; 2445-2459; 2466-2469, 'X', 2471-2474; 2504-2508, 'X',
A/Cross-references: UNIPARC:UPI0000173C15; UNIPARC:UPI0000173C16; UNIPARC:UPI0000173C17;
CIC: UNIPARC:UPI0000173C1D; UNIPARC:UPI0000173C1E; UNIPARC:UPI0000173C1F; UNIPARC:UPI000
10000173C25; UNIPARC:UPI0000173C26; UNIPARC:UPI0000173C27
R/Zanussi, S.; Dollana, R.; Segat, D.; Bonaldo, P.; Colomacchi, A.
J. Biol. Chem. 267, 24082-24089, 1992
A/Reference number: S28776; MUID:93054780; PMID:13399440
A/Accession: A57083
A/Molecule type: DNA
A/Residues: 310-328 <ZAN>
A/Cross-references: UNIPARC:UPI0000173C28
A/Accession: S28776
A/Molecule type: mRNA
A/Residues: 32-126, 'AK', 129-136, 'U', 138-236 <ZAZ>
A/Cross-references: UNIPARC:UPI0000173C28; GB:S49432; NID:G260296; PIDN:AA24261.1; PID:
R/Chu, M.L.; Mann, K.; Deutmann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.;
Eur. J. Biochem. 168, 309-317, 1987
A/Title: Characterization of three constituent chains of collagen type VI by peptide seq
A/Reference number: S00126; MUID:88029444; PMID:3665927
A/Accession: S00245
A/Molecule type: mRNA, protein
A/Residues: 2024-2046; 2092-2156, 'R', 2203-2208, 'X', 2210-2211, 'X', 2213-2227; 2228-2251; 2314
A/Cross-references: UNIPARC:UPI000016A705; UNIPARC:UPI0000173C29; UNIPARC:UPI0000173C2A;
ID:G1335034
A/Note: The mRNA portion of the sequence corresponds to residues 2092-2157
R/Chu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutmann, R.; Timpl, R.
J. Biol. Chem. 263, 18601-18606, 1988
A/Title: Amino acid sequence of the triple-helical domain of human collagen type VI.
A/Reference number: A31952; MUID:89066644; PMID:3198591
A/Accession: C31952
A/Molecule type: mRNA
A/Residues: 2038-2373 <CH4>
A/Cross-references: UNIPARC:UPI0000173C2E; GB:J04211; GB:M20778
A/Note: parts of this sequence were determined by protein sequencing
R/Welch, D.; Mattei, M.G.; Passage, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deutz
Am. J. Hum. Genet. 42, 435-445, 1988
A/Title: Cloning and chromosomal localization of human genes encoding the three chains o
A/Reference number: A29848; MUID:88161046; PMID:3348212
A/Accession: C29848
A/Molecule type: mRNA
A/Residues: 2092-2151 <MEI>
A/Cross-references: UNIPARC:UPI000016A720; GB:M27449; NID:G291919; PIDN:AA52057.1; PID:
A/Note: part of this sequence was determined by protein sequencing
R/Jander, R.; Rautenberg, J.; Glanville, R.W.
Eur. J. Biochem. 133, 39-46, 1983
A/Title: Further characterization of the three polypeptide chains of bovine and human sh


```

Db      75 ThrAlaPheLeuAaSerLeuLeuArgAsnMetAspIleGlyProGlnGlnThrGlnVal 94
Qy      124 AGCTGTCGACAGTCGGAGTCGCGCATACACCGAGTCCCTTGCGGCACAGACAGCTCG 183
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      95 GlyIleValGlnIrrGlyGlnThrValAlaHisGluPheTyrLeuAaThrTyrSerThr 114
Qy      184 GGTGAGCTGCCCGAGATCGCGTGCCTTGCCTCCGACGATGGT--GACACCCAC 240
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      115 ThrGluGlnValMetAspAlaAlaLeuArgIleArgGlnArgGlyGlyThrGlnThrMet 134
Qy      241 ACTGGCCGCGGCGCTGTCTATGCCAAGAGACCTGTTTGGTGCAGATCGACGCGCCGG 300
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      135 ThrAlaLeuGlyIleAspThrAlaArgGluGlnAlaPheThrGlnAlaHisGlyAlaArg 154
Qy      301 CCAGGAGTGCACCAAGTCTGTGTGTGTGACAGATGGCGGCTCCAGGACCTGTGGGC 360
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      155 ArgGlyValGlnIrrValMetValIleValThrAspGlyGlnSerHisAsp-----Asn 172
Qy      361 CCCCCATGACAGAGCTC-----AAGACCTGGGCGCTACCCGTGTTTCATTGTC 408
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      173 TyrArgLeuGlnGlnValIleAspIrrGlyAspGluAsnIleGlnArgPheAlaIle 192
Qy      409 -----AGACCGGCGGAGGCAACTTCCTG-----GAGCTG 438
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      193 AlaIleLeuGlySerTyrSerArgIrrGlyAsnLeuSerThrGluIrrPheValGlnGluIrr 212
Qy      439 TCAGCCGCTGCTCAGCCCTGCGAGAGACCTGCACTTGTGACGTGGATGAC--- 495
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      213 LysSerIleAlaSerIrrPheProThrGluIrr-----HisPhePheAsnValSerAspGlu 230
Qy      496 -----CTGCACATCATTTGTCACAGACTGAGGGGCTCCATTCTCGCG 537
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      231 LeuAlaLeuValThrIleValGlnIrrLeuGlyGluArgIlePheAla 246

RESULT 13
type XII collagen alpha-1 chain - eastern newt (fragment)
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: I51027
R:Wei, Y.; Yang, E.V.; Klatz, K.P.; Tassava, R.A.
Dev. Biol. 168, 503-513, 1995
A:Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII colla
A:Reference number: I51027; MUID:95246925; PMID:7729585
A:Accession: I51027
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-929 <WEI>
A:Cross-references: UNIPROT:Q91145, UNIPARC:UPI000012602E, EMBL:U19494, NID:9632647, PID
F:155-236/Domain: fibronectin type III repeat homology <3FR>
F:631-795/Domain: von Willebrand factor type A repeat homology <VWA3>

Alignment Scores:
Pred. No.: 5,07e-08 Length: 929
Score: 197.50 Matches: 50
Percent Similarity: 48.6% Conservative: 36
Best Local Similarity: 28.2% Mismatches: 90
Query Match: 18.8% Indels: 1
DB: 2 Gaps: 1

US-10-699-035A-1 (1-537) x I51027 (1-929)
Qy      4 GACCTGATGTCCTGCTGACAGCTCAGCCAGCTCTCTCACTACAGAGTTCTCCGGGT 63
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      633 AspIleValIrrLeuValAspIrrGlySerTrrPssrIleGlyArgProAsnPheIrrIrrVal 652
Qy      64 CGGAGTTTGGGGGAGCTGTGCTCCTCAGCTCCCTGGGACCGGGGCGCTGCGTGC 123
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      653 ArgAsnPheIrrSerArgValIrrValIrrPheAspIrrIrrGlySerAspArgValGlnIrr 672
Qy      124 AGCTGTCGACAGTCGGAGCTGCGCATACAGAGTTCCTTGGCGGACAGACAGCTCG 183
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      673 AlaAlaSerGlnTrrSerGlyAspProArgThrGlnTrrGlnLeuAaThrThrIrrSerThr 692

```

```

Qy      184 GGTGAGCTGCCAGATGCGGTGCGTCTTCTGCGCAGCGATGGTGAACCCACACT 243
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      693 LysIrrSerLeuMetAspAlaValAlaAsnLeuProTrrIrrGlyIrrAsnThrAsnThr 712
Qy      244 GCGCTGCGCTGCTCTATGCGACCAAGACAGCTTTTGTGAAGATCAGTGGCCCGGCA 303
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      713 GlySerAlaLeuIrrPheIrrLeuGlnAsnAsnAsnHeArgProGlyValIrrGlyMetArgGlu 732
Qy      304 GGGGTGCCCAAGTGTGTGTGTGTGTGACAGATGGCGGCTCCAGCAGCCCTGCGGCCCC 363
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      733 LysAlaArgIrrIrrIrrAlaIrrLeuThrAspIrrIrrIrrSerGlnAspAspIrrIrrVal 752
Qy      364 CCATGACAGAGCTCAAGAGACTGGGCGTCAACCGTTCATTGTTCAGACACCGCGGAGCC 423
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      753 ProSerIrrArgTrrAlaAspGluGlyIrrLeuLeuTrrAlaValIrrGlyIrrIrrAsnAla 772
Qy      424 AACTCTCGAGCTGTACAGCGCTGCTGCTGACCGCTCGCGGAGAACCTGACCTTGTG 483
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      773 AspGluAsnGlnIrrLeuIrrGlnIrrIrrIrrAspProAspGlnLeuTrrIrrMetTrrAsnVal 792
Qy      484 ---GACGTGATGACCTGACATCATGTCACAGAGCTGAGGGGCTCCATT 531
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      793 AlaAspPheSerLeuLeuThrAsnIrrIrrValAsnAspLeuThrGluAsnVal 809

RESULT 14
integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Cross-references: UNIPROT:P56199, UNIPARC:UPI0001209EB
A:Experimental source: hepatoblastoma cell line HepG2
A:Note: sequence extracted from NCBI backbone (NCBI:P124326)
F:142-317/Domain: von Willebrand factor type A repeat homology <VWA1>

Alignment Scores:
Pred. No.: 5,49e-08 Length: 1151
Score: 197.00 Matches: 65
Percent Similarity: 47.9% Conservative: 28
Best Local Similarity: 33.5% Mismatches: 81
Query Match: 18.8% Indels: 20
DB: 2 Gaps: 7

US-10-699-035A-1 (1-537) x A45226 (1-1151)
Qy      4 GACCTGATGTCCTGCTGACAGCTCAGCCAGCTCTCTCACTACAGAGTTCTCCGGGT 63
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      144 AspIrrValIrrIrrValIrrLeuAspGlySerAsnSerIrr-----TrrProTrrAspSerVal 161
Qy      64 CGGAGTTTGGGGGAGCTGTGCTCCTCAGCTCCCTCGGACCGGGGCGCTGCGTGC 123
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      162 ThrAlaPheLeuAaAspLeuIrrLeuIrrIrrArgMetAspIrrIrrGlyProIrrGlnThrGlnVal 181
Qy      124 AGCTGTCGACAGTCGGAGTCGCGCATACACAGAGTTCCTTGGCGGACAGACAGCTCG 183
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      182 GlyIrrValIrrIrrIrrGlyIrrAsnValThrIrrIrrIrrPheAsnLeuAaAsnIrrTrrSerSer 201
Qy      184 GGTGAGCTGCCAGAGTGGTGTGCTGCTTTCGCGGACCGCATGGT--GACACCCAC 240
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      202 ThrGluGlnValIrrValIrrAlaAlaIrrIrrIrrIrrValIrrGlnArgGlyIrrArgGlnThrMet 221
Qy      241 ACTGCGCTGCGGCGCTGTATGCGCAAGAGAACAGCTTGTGTAAGATCAGTGGCGCGG 300
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      222 ThrAlaLeuGlyIrrThrAspThrAlaArgIrrGlnIrrIrrPheThrGlnAlaArgIrrAlaArg 241

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: February 13, 2006, 13:13:13 ; Search time 27.3147 Seconds
(without alignments)
2774.097 Million cell updates/sec

Title: US-10-699-035A-1

Perfect score: 1049
Sequence: 1 999gaccctgacgtctcctcgt.....tgaggggctccattcgcgcg 537

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-O=/abs/ABSSREP.spool/US10699035/runat_13022006_062441_25416/app.query.fasta_1
-Db=uniprot -OPMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs902p
-USER=US10699035 @CCN 1.1 580 @runat_13022006_062441_25416 -NCPU=6 -ICPU=3
-NM MAP -NEG SCORES=0 -WALT -DSPLITLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
Database : Uniprot 05.80:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	902	86.0	445	2	06PCB0 HUMAN
2	766	73.0	415	2	08R225 MOUSE
3	762	72.6	415	2	0923K3 MOUSE
4	762	72.6	415	2	08C007 MOUSE
5	754	71.9	415	2	0642A6 RAT
6	395	37.7	505	2	04SCD1 TETNG
7	273	26.0	2225	2	04SXK3 TETNG
8	272.5	26.0	1723	2	04SPD2 TETNG
9	268	25.5	1888	1	COEAI CHICK
10	267.5	25.5	1259	1	QATP12 TETNG
11	265	25.3	624	1	MATN4 MOUSE
12	263.5	25.1	3124	1	COCAI CHICK
13	263	25.1	1557	2	04SH63 TETNG
14	261.5	24.9	2884	2	05VYK2 HUMAN
15	261.5	24.9	3063	1	COCAI HUMAN
16	261.5	24.9	3063	2	05VYK1 HUMAN

17	259	24.7	622	1	MATN4 HUMAN	095460 homo sapien
18	257.5	24.5	517	2	043853 HUMAN	043853 homo sapien
19	254.5	24.3	637	2	081VX1 HUMAN	081VX1 homo sapien
20	254.5	24.3	1284	2	04VXQ5 HUMAN	04VXQ5 homo sapien
21	254.5	24.3	1284	2	06P159 HUMAN	06P159 homo sapien
22	254.5	24.3	1297	2	04VXQ4 HUMAN	04VXQ4 homo sapien
23	254.5	24.3	1329	1	KI510 HUMAN	09P218 homo sapien
24	254.5	24.3	1796	1	COEAI HUMAN	005707 homo sapien
25	251	23.9	688	2	04TOK3 TETNG	04TOK3 tetraodon n
26	249	23.7	839	2	06P3N7 XENTR	06P3N7 xenopus tro
27	248.5	23.7	1117	2	04RXN8 TETNG	04RXN8 tetraodon n
28	248	23.6	644	2	05NUJ1 BRARE	05NUJ1 brachydanio
29	248	23.6	821	2	06PYX2 BRARE	06PYX2 brachydanio
30	247.5	23.6	944	2	05NUJ5 BRARE	05NUJ5 brachydanio
31	247.5	23.6	493	1	MATN1 CHICK	P05099 gallus galli
32	247.5	23.6	1797	1	COEAI MOUSE	080X93 gallus galli
33	246.5	23.5	451	2	08N2G3 HUMAN	08N2G3 homo sapien
34	246.5	23.5	956	1	MATN2 HUMAN	000339 homo sapien
35	246	23.5	3119	1	COCAI MOUSE	060847 mus musculu
36	245.5	23.4	534	2	04GOW3 HUMAN	04GOW3 homo sapien
37	245.5	23.4	955	2	05R9N1 PONPY	05R9N1 pongo pygma
38	245	23.4	685	2	05NUJ2 BRARE	05NUJ2 brachydanio
39	242	23.1	490	2	07SYT5 XENTLA	07SYT5 xenopus lae
40	242	23.1	726	2	05NUJ4 BRARE	05NUJ4 brachydanio
41	241.5	23.0	647	2	04S2X7 TETNG	04S2X7 tetraodon n
42	240.5	22.9	500	2	08OVN5 MOUSE	08OVN5 mus musculu
43	240	22.9	261	2	05CZQ6 BRARE	05CZQ6 brachydanio
44	239.5	22.8	500	1	MATN1 MOUSE	P51942 mus musculu
45	239.5	22.8	1450	2	Q4RP14 TETNG	Q4RP14 tetraodon n

ALIGNMENTS

RESULT 1

06PCB0 HUMAN
ID 06PCB0; HUMAN PRELIMINARY; PRT; 445 AA.
AC 06PCB0;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein, isoform 1 (WARP).
GN Name=WARP; ORFNames=RP4-758018.11-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marcovina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Miliady S.J.,
RA Bosak S.A., Weisman P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Placenta;
RA Strausberg R.;
RN Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
[3]
RP NUCLEOTIDE SEQUENCE.
RA Hall R.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC059409; AAH59409.1; -; mRNA.
DR EMBL; AL391244; CA122657.1; -; Genomic DNA.
DR EMBL; ENSG00000179403; Homo sapiens.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00041; FN3; 2.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50853; FN3; 2.
DR PROSITE; PS50234; VWF_A; 1.
SQ SEQUENCE 445 AA; 46804 MW; D9FBDE9C0A4DEBDF CRC64;

Alignment Scores:
Pred. No.: 1 03e-59 Length: 445
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 2 Gaps: 0

US-10-699-035A-1 (1-537) x Q6P8B0_HUMAN (1-445)

QY 1 GGGGACCTGATGTTCTCTGCTGACAGCTCAGCCAGCTGCTCTACAGAGTTCTCCCGG 60
DB 33 G1YAAPLeuMet.PheLeuLeuMet.SerSerAlaSerValSerHisIleTyrGluPheSerArg 52
QY 61 GTTCGGAGTTTGTGGGAGAGCTGGTGGTCCAGTCCCGGCGGAGCGGCGGCGTGGCT 120
DB 53 VALArgGluPheValGluGluLeuValAlaProLeuProLeuGluTyrThrIleValLeuVal 72
QY 121 GCCAGTCTGTGACAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGC 180
DB 73 AlAserIleValAlaHisValGluSerArgProTyrThrGluPheProPheGluGluHisSer 92
QY 181 TCGGGTGAGGCTGCCCGAGAGTGGGTGCTTCTGCCAGCGCATGGGTGACACCCAC 240
DB 93 SerGluGluAlaAlaGluAlaSerAlaValAlaGluAlaSerAlaGluArgMetGluYlePThrHis 112
QY 241 ACTGGCTGGCGCTGGTCTATGCGCAAGAAAGCTTTGCTGGAAGCATCAGTGGCCGG 300
DB 113 ThrGluLeuAlaLeuValTyrAlaTyrGluGluLeuPheAlaGluAlaSerGluYalaArg 132
QY 301 CCAGGGGTCGCCAAAGTGTGCTGTTGGTGGAGATGGGCGGCTCCAGCAGCCCTGGAGC 360
DB 133 ProGluYalProIleValLeuValAlaTyrValThrAspGluGluSerSerAspProValGlu 152
QY 361 CCCCCCATGACAGAGCTCAGAGCTGCGGCGTCCAGCTGTTCTATGTCAGACCGCCGA 420
DB 153 ProprometGluGluLeuLeuSerAlaAlaAlaSerAlaProAlaGluYleHisIleHisPhe 172
QY 421 GGCACTTCTCTGAGCTGTACGCCGCTGCTCAGCCCGCTCCGGAAGAGACCTGCACTTT 480
DB 173 G1YAsPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluYleHisIleHisPhe 192
QY 481 GTGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 534
DB 193 ValAspValAspAspLeuHisIleIleValGluGluLeuGluYleSerIleLeu 210

RESULT 2
Q8R2Z5 MOUSE PRELIMINARY; PRT; 415 AA.
AC Q8R2Z5
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein (Mus musculus 11 days
DE pregnant adult female ovary and uterus cDNA, RIKEN full-length
DE enriched library, clone:5031410123 product:VON WILLEBRAND FACTOR A-
DE RELATED PROTEIN homolog).
CN Name=Wval;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II, and FVB/N;
RC TISSUE=Colon, and
RX Mammary tumor metastasized to lung. Tumor arose spontaneously;
MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshitsuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX NIH MGC Project;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning".
RL Meth. Enzymol. 303:19-44 (1999).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi T., Fukuda S.,
RA Saito K., Izawa M., Nishi K., Kiyosawa H., Kando S., Yamanaoka I.,
RA Aizawa T., Okazaki Y., Gojobori T., Bono H., Kanukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Macnau Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shihada Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690(2001).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RA The FANTOM Consortium.
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RC MEDLINE=20493974; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada K.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RC Adachi U., Aizawa K., Akimura T., Aikawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Kurita H., Kawai J., Kojima Y., Kondo S., Kono H., Konda M., Koya S.,
 RA Kuribara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ono M., Onisato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Taya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=FVB/N; TISSUE=Colon;
 RA Director MGC Project.
 RL Submitted (Jul-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC026919; AAH26919.1; -; mRNA.
 DR EMBL; AK077240; BAC36703.1; -; mRNA.
 DR EMBL; BC036166; AAH36166.1; -; mRNA.
 DR HSSP; P18614; IMHP.
 DR Ensembl; ENSMUSG0000042116; Mus musculus.
 DR MGI; MGI:2179729; Wval.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR002035; WVF_A.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00092; WVA; 1.
 DR PRINTS; SM00453; VMPADOMAIN.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00327; WVA; 1.
 DR PROSITE; PS50853; FN3; 2.
 DR PROSITE; PS50834; WVA; 1.
 DR PROSITE; PS50834; WVA; 1.
 SQ SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;
 Alignment Scores: 1.95e-49 Length: 415
 Pred. No.: 766.00 Matches: 150
 Score:

Percent Similarity: 91.5% Conservative: 12
 Best Local Similarity: 84.7% Mismatches: 15
 Query Match: 73.0% Indels: 0
 DB: 2 Gaps: 0
 US-10-699-035a-1 (1-537) x QBR25_MOUSE (1-415)
 QY 1 GGGGACCTGATGTTCTCTGAGCAGCTCAAGCAGCTCTCTCACTACGAGTTCTCCCG 60
 DB 33 GYAAPLeuLeuPheLeuLeuLeuPheSerSerSerValSerHisTyrGluPheSerArg 52
 QY 61 GTTCGGAGTTTGTGGGGCAGCTGTGCTCCACTGCCCCCGGACCGGGCCCTGGCT 120
 DB 53 VALAGGluPheValGlyGluLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 72
 QY 121 GCCAGTCTGGGACAGTGGGGCAGTGGGCATACACGAGTTCCCTCGGACGACAGC 180
 DB 73 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheArgGlnTyrSer 92
 QY 181 TCGGCTGAGGCTGCCAGGATGCGGTGCTTCTGCGCCAGCGGATGGTGACACCCAC 240
 DB 93 SerGlyGlnAlaIleGluAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 112
 QY 241 ACTGCGCTGGCGCTGTCTATGCAAGAACAGCTGTTTCTGAAGCATCAGTGGCCGG 300
 DB 113 ThrGlyLeuAlaLeuAlaTyrAlaGlyGluLeuPheAlaGluGluAlaGlyAlaArg 132
 QY 301 CCAAGGCGTCCCAAGTCTGCTGGTGGTGACAGATGGCGGCTCCAGGACCTGGGGC 360
 DB 133 ProGlyValProValValLeuValTyrValThrAspGlyGlySerSerAspProValGly 152
 QY 361 CCCCCATGCAAGAGCTCAAGAGCTGGCGCTGACCGTGTTCATTGTTCAGACCGGCGA 420
 DB 153 ProprometGlnGluLeuLeuAspLeuGlyValThrIlePheIleValSerThrGlyArg 172
 QY 421 GGCACCTTCCTGAGACTGTACGCCCTGCTTACAGCCCTGCGGAGAGAGACTGCACTT 480
 DB 173 GlyAsnLeuLeuGluLeuLeuAlaAlaSerAlaProAlaGluLeuHisLeuHisPhe 192
 QY 481 GTTGAGCTGATGATGACCTGCATCTTCCTCAAGAGCTGAGGGGCTCCATT 531
 DB 193 ValAspValAspAspLeuProIleAlaArgGluLeuArgGlySerIle 209
 RESULT 3
 Q923K3_MOUSE PRELIMINARY; PRT; 415 AA.
 ID Q923K3_MOUSE PRELIMINARY; PRT; 415 AA.
 AC Q923K3;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Von Willebrand factor A-related protein.
 GN Name=Wval; Synonyms=4932416A1Rik, Warp;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC MEDLINE=22057805; PubMed=12062410; DOI=10.1016/S0014-5793(02)02579-6;
 RA Fitzgerald J., Ting S.T., Bateman J.F.,
 RT "WARP a new member of the von Willebrand factor A-domain superfamily
 RT of extracellular matrix proteins."
 RL FEBS Lett. 517:61-66(2002).
 DR EMBL; AY030094; AAK38350.1; -; mRNA.
 DR HSSP; P18614; IMHP.
 DR Ensembl; ENSMUSG0000042116; Mus musculus.
 DR MGI; MGI:2179729; 4932416A1Rik.
 DR MGI; MGI:2179729; Wval.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR002035; WVF_A.
 DR Pfam; PF00041; fn3; 2.

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK030019; BAC26739.1; -, mRNA.
DR HSSP: P18614; 1MHP.
DR Ensembl: ENSMUSG0000042116; Mus musculus.
DR MGI: MGI:2179729; 4932416A1.Rik.
DR MGI: MGI:2179729; Vwa1.
DR GO: GO:0005615; C:extracellular space; TAS.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR02035; VWF_A.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00092; VWA; 1.
DR PRINTS: PR00453; VWFADOMAIN.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS00853; FN3; 2.
DR PROSITE: PS0234; VWF_A; 1.
DR PROSITE: PS0234; VWF_A; 1.
SO SEQUENCE 415 AA; 44681 MW; CE137963B76834FE CRC64;

Alignment Scores:
Pred. No.: 3,91e-49 Length: 415
Score: 762.00 Matches: 149
Percent Similarity: 91.0% Conservative: 12
Best Local Similarity: 84.2% Mismatches: 16
Query Match: 72.6% Indels: 0
DB: Gaps: 0

US-10-699-035a-1 (1-537) x Q8CQ07_MOUSE (1-415)

QY 1 GGGGACCTGATGTTCTGTCGACAGTCCAGGCTGTCCTCATTACAGAGTTCTCCCG 60
DB 33 G1yAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerIstIryGluPheSerArg 52
QY 61 GTTCCGAGATTGTTGGGGAGAGCTGTCGTCCTCCAGTCCCGCCCGGAGCCGCGCTGCT 120
DB 53 ValArgGluPheValG1yGlnLeuValAlaThrMetSerPheG1yProG1yAlaLeuArg 72
QY 121 GCCAGTCTGTCGACGTCGAGTCGAGTCGAGCATACACCGAGTTCCTTCGCGCCAGCAGC 180
DB 73 AlaSerLeuValHisValG1ySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92
QY 181 TGGGGTGAAGGCTCCAGAGTGGCGGCTGTCCTTCGCCAGGAGGTCGTCAGACCCAC 240
DB 93 SerG1yGlnAlaIleGlnAspAlaIleArgValAlaProGlnArgMetG1yAspIlnraen 112
QY 241 ACTGGCCTGGCCTGTCATGTCGACAGAAAGAGCTTTGTCGAAACATCAGGTCGCCG 300
DB 113 ThrG1yLeuAlaLeuAlaIlyrAlaIySg1yGlnLeuPheAlaGlnIuAlaG1yAlaArg 132
QY 301 CCAGGGGTCGCCAAGTGTGTCGTGTCGTCGATGCGCGCTCCAGCAACCTGTGGC 360
DB 133 ProG1yValProIyValLeuValItrValThrAspG1yG1ySerSerAspProValG1y 152
QY 361 CCCCCCATGAGAGCTCAAGACCTGGCGCTCACCGTGTATTCATTTCACGACCGGCCA 420
DB 153 ProProMetGlnG1yLeuLeuAspLeuG1yAlaThrIlePheIleValSerThrG1yArg 172
QY 421 GCGAATTCCTGAGAGTGTGTCAGCCGTCGTCGACCCCTGCGCAGAGACCTGACTT 480
DB 173 G1yAspLeuLeuGlnLeuLeuAlaAlaAlaSerAlaProIaGlnIuIySleuHisPhe 192
QY 481 GTGAGCTGATGACCTGACATCATGTCAGAGCTGAGGGGCTTCATT 531
DB 193 ValAspValAspAspLeuProIleIleAlaArgGluLeuArg1ySerIle 209

RESULT 5
Q642A6_RAT
ID Q642A6_RAT PRELIMINARY; PRT; 415 AA.
AC Q642A6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Von Willebrand factor A domain-related protein.
GN Name=RGD1311476_Predicted;

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schlier G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshlyuk S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Mizny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalys D.E.,
RA Scheraga A., Schein J.E., Jones S.J.M., Maira M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RG NIH MGC Project;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC081983; AA081983.1; -, mRNA.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR02035; VWF_A.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00092; VWA; 1.
DR PRINTS: PR00453; VWFADOMAIN.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS00853; FN3; 2.
DR PROSITE: PS0234; VWF_A; 1.
SO SEQUENCE 415 AA; 44832 MW; B3A0A0268DB06D4E CRC64;

Alignment Scores:
Pred. No.: 1,58e-48 Length: 415
Score: 754.00 Matches: 149
Percent Similarity: 90.4% Conservative: 12
Best Local Similarity: 83.7% Mismatches: 17
Query Match: 71.9% Indels: 0
DB: Gaps: 0

US-10-699-035a-1 (1-537) x Q642A6_RAT (1-415)

QY 1 GGGGACCTGATGTTCTGTCGACAGTCCAGGCTGTCCTCATTACAGAGTTCTCCCG 60
DB 33 G1yAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisIryGluPheSerArg 52
QY 61 GTTCCGAGATTGTTGGGGAGAGCTGTCGTCCTCCAGTCCCGCCCGGAGCCGCGCTGCT 120
DB 53 ValArgGluPheValG1yGlnLeuValAlaThrMetSerPheG1yProG1yAlaLeuArg 72
QY 121 GCCAGTCTGTCGACGTCGAGTCGAGTCGAGCATACACCGAGTTCCTTCGCGCCAGCAGC 180
DB 73 AlaSerLeuValHisValG1ySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92
QY 181 TGGGGTGAAGGCTCCAGAGTGGCGGCTGTCCTTCGCCAGGAGGTCGTCAGACCCAC 240
DB 93 SerG1yGlnAlaIleGlnAspAlaIleArgValAlaProGlnArgMetG1yAspIlnraen 112
QY 241 ACTGGCCTGGCCTGTCATGTCGACAGAAAGAGCTTTGTCGAAACATCAGGTCGCCG 300

CC	preliminary data.
DR	EMBL; CAAB01012445; CAF9668.1; -, Genomic_DNA.
DR	InterPro; IPR008160; Collagen.
DR	InterPro; IPR003191; FN III.
DR	InterPro; IPR003191; laminin G_TSP_N.
DR	InterPro; IPR001220; lectin_1egB.
DR	InterPro; IPR002035; WVF_A.
DR	Pfam; PF01391; Collagen; 2.
DR	Pfam; PF00041; fn3; 8.
DR	Pfam; PF00092; WVA; 3.
DR	PRINTS; PR00453; VWFDOMAIN.
DR	SMART; SM00060; FN3; 9.
DR	SMART; SM00210; TSPN; 1.
DR	SMART; SM00327; WVA; 3.
DR	PROSITE; PSS0853; FN3; 9.
DR	PROSITE; PSS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
DR	PROSITE; PSS0234; WVA; 3.
FT	Collagen; Extracellular matrix; structural protein.
EQ	NON_TER 1 1
SEQUENCE	2225 AA; 239674 MW; EC6545B60A778BD9 CRC64

FT CARBOHYD 138 138 N-linked (GlcNAc...) (Potential)
 FT CARBOHYD 1398 1398 N-linked (GlcNAc...) (Potential)
 FT TURN 1855 1855
 FT TURN 1859 1861
 FT HELIX 1862 1880
 FT TURN 1881 1882
 SQ SEQUENCE 1888 AA; 202668 MW; 39915BB9F46DD873 CRC64;

Alignment Scores:

Pred. No.: 9, 77e-12 Length: 1888
 Score: 268.00 Matches: 64
 Percent Similarity: 50.5% Conservative: 34
 Best Local Similarity: 33.0% Mismatches: 80
 Query Match: 25.5% Indels: 16
 DB: 1 Gaps: 2

US-10-699-035A-1 (1-537) x COEAL_CHICK (1-1888)

QY 4 GACCTGATGTCCTGCTGAGACCTGACCGAGCTCTCACTACGAGTTCTCCGGGTT 63
 DB 158 AapilValIleValValaspIysertIpsertIleGlyArgpheaSpheargLeuVal 177
 QY 64 CGGAGTTGTGGGAGCTGTGCTCCACTGCCCCCTGGGACCGGGCCCTGCTGCC 123
 DB 178 ArgLeupheLeugluuAsnLeuValSerAlaPheaSnValGlyserGluYthrArgVal 197
 QY 124 AGCTGTGTCACGTGGGACGTGGCCATACCGAGTCCCTTCGGCCAGACAGCTCG 183
 DB 198 GlyLeuAlaGlnTyrserIysArgProAlaGlnIleGluThrIleuAsnAlaTyrGlyThr 217
 QY 184 GGTGAGCTGCCAGAGTCCGCTGCTCTTCTGCCAGCGATGGTGACACCCACT 243
 DB 218 LysAspAlaValIleuAspAlaValArgAsnLeuProTyrIleGlyYasnThrLeuThr 237
 QY 244 GGCCTGGCGCTGCTATGCAAGAAACAGCTGTTGTCGAGACATCAGTCCCGGCA 303
 DB 238 GlyLeuAlaLeuThrTyrIleLeuGluAsnSerPheLysProGluAlaGlyAlaArgPro 257
 QY 304 GGGGATGCCAAGTGTGCTGAGTGCAGATGAGCGGCTCCAGACCCCTGAGCGGCC 363
 DB 258 GlyValSerLysIleGlyIleuIleThrAspGlyLysSerGlnAspValIlePro 277
 QY 364 CCCATGAGAGCTCAAGACCTGGCGTCAACCGTTCATTGTCAGACCGGCGAGGC 423
 DB 278 ProAlaLysAsnLeuArgAspAlaGlyIleGluLeuPheAlaIleGlyValLysAsnAla 297
 QY 424 AACTTCCTGAGCTGTCAAGCCGCTGCTCAAGCCCTGCGAGAGACCTGCATTGTG 483
 DB 298 AapilLeasngluLeuYsgluIleAlaSerGluProAspSerThrIleValTyrAsnVal 317
 QY 484 ---GACGTGATGATGACCTGCATCATTCATTCACAA----- 513
 DB 318 AlaAspPheAsnPheMetAsnSerIleValGluGlyLeuThrArgThrValCysSerArg 337
 QY 514 -----GAGCTGAGGGGCTCCATTCTCGCG 537
 DB 338 ValGluGluGlnGluLysGluIleYsglyThrIleAlaAla 351

RESULT 10

QARP12_TETNG PRELIMINARY; PRT; 1259 AA.
 AC QARP12;
 DT 13-SBP-2005 (TREMBLrel. 31, Last sequence update)
 DT 13-SBP-2005 (TREMBLrel. 31, Last annotation update)
 DE Chromosome 10 (SCAF15009, whole genome shotgun sequence.
 GN ORFNames=GSTENG00031322001;
 OC Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 NCBI_TaxId=99883;

RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Strange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicoud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Dubin C., Castellico L., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Catalicio L., Poulain J., De Bernardis V.,
 RA Cruaud C., Duprat S., Brottier P., Coullencan J.P., Gouzy J.,
 RA Parra G., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Landel V., Schachter V., Queller F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Croillins H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -! CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; CAEO1015009; CAG09870.1; -; Genomic DNA.
 SQ SEQUENCE 1259 AA; 138606 MW; 844C01B6FE3D0E5D CRC64;

Alignment Scores:

Pred. No.: 1, 03e-11 Length: 1259
 Score: 267.50 Matches: 65
 Percent Similarity: 52.5% Conservative: 28
 Best Local Similarity: 36.7% Mismatches: 83
 Query Match: 25.5% Indels: 1
 DB: 2 Gaps: 1

US-10-699-035A-1 (1-537) x QARP12_TETNG (1-1259)

QY 4 GACCTGATGTCCTGCTGAGACCTGACCGAGCTCTCACTACGAGTTCTCCGGGTT 63
 DB 395 AapilValIleValValaspIysertIpsertIleGlyLeuGlnAsnheAlaVal 414
 QY 64 CGGAGTTGTGGGAGCTGTGCTCCACTGCCCCCTGGGACCGGGCCCTGCTGCC 123
 DB 415 ArgAlaPheLeuGluValIleuValAsnSerPheAspIleGlyProSerIysValGlnIle 434
 QY 124 AGCTGTGTCACGTGGGACGTGGCCATACCGAGTCCCTTCGGCCAGACAGCTCG 183
 DB 435 SerLeuValGlnTyrserArgAspProHlsThrGluPheAlaLeuAsnThrIshIsaSp 454
 QY 184 GGTGAGAGCTGCAGAGTCCGCTGCTCTGCCAGCGATGGTGACACCCACT 243
 DB 455 IleAsnAlaValAlaArgAlaValArgThrPheProTyrArgGlyGlyserThrAsnThr 474
 QY 244 GGCCTGGCGCTGCTATGCAAGAAACAGCTGTTGTCGAGACATCAGTCCCGGCA 303
 DB 475 GlyLysAlaMetLysTyrValLysAspLysIlePheValAlaSerArgGlyAlaArgGln 494
 QY 304 GGGGATGCCAAGTGTGCTGAGTGCAGATGAGCGGCTCCAGACCCCTGAGCGGCC 363
 DB 495 AsnValProArgValMetValLeuIleThrAspGlyLysSerSerAspSerPheLysAsp 514
 QY 364 CCCATGAGAGCTCAAGACCTGGCGTCAACCGTTCATTGTCAGACCGGCGAGGC 423
 DB 515 AlaAlaThrAsnLeuArgAsnIleAspValGluIlePheAlaValGlyValLysAspAla 534
 QY 424 AACTTCCTGAGCTGTCAAGCCGCTGCTCAAGCCCTGCGAGAGACCTGCATTGTG 483
 DB 535 ValArgSerGluLeuGluAlaIleAlaAsnPropAlaAspAsnIleValPheGluVal 554
 QY 484 ---GACGTGATGATGACCTGCATTCATTCACAAAGCTGAGGGGCTCCATT 531
 DB 555 GluAspPheAspAlaPheGlnArgIleSerIysGluLeuThrGlnSerIle 571

RESULT 11	MATN4_MOUSE	STANDARD:	PRT:	624 AA.
ID	MATN4_MOUSE	089029;	Q9QWS3;	
AC		16-OCT-2001 (Rel. 40, Created)		
DT		16-OCT-2001 (Rel. 40, Last sequence update)		
DT		10-MAY-2005 (Rel. 47, Last annotation update)		
DE		Matrilin-4 precursor (MAT-4).		
GN		Name=Matn4;		
OS		Mus musculus (Mouse).		
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC		Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;		
OC		Muroidea; Muridae; Murinae; Mus.		
OX		NCBI_TaxId=10090;		
RN		[1]		
RP		NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT), AND VARIANTS.		
RC		STRAIN=C57BL/6J, and CD-1; TISSUE=Fetal;		
RX		MEDLINE=98442849; PubMed=9771906; DOI=10.1016/S0014-5793(98)01111-9;		
RA		Wagener R., Kobbe B., Paulsson M.,		
RT		"Matrilin-4, a new member of the matrilin family of extracellular		
RL		matrix proteins."		
RL		FEBS Lett. 436:123-127(1998).		
RN		[2]		
RP		NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM LONG).		
RC		STRAIN=C57BL/6J; TISSUE=Mammary gland;		
RX		MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA		Steinberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA		Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA		Aleischl S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA		Hopkins R.F., Jordan H., Moore T., Max J., Wang J., Hsieh F.,		
RA		Diachenko L., Marinina K., Farmer A.A., Rubin G.M., Hong L.,		
RA		Stapleton K., Soares M.B., Donald M.F., Cabavant T.L., Scheetz T.E.,		
RA		Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA		Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA		Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA		Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,		
RA		Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA		Falley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,		
RA		Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA		Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA		Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,		
RA		Butterfield V.S.N., Krzywinski M.T., Skalska U., Smalhus D.E.,		
RA		Scherer A., Schein J.E., Jones S.J.W., Marra M.A.;		
RT		"Generation and initial analysis of more than 15,000 full-length human		
RT		and mouse cDNA sequences."		
RL		Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN		[3]		
RP		NUCLEOTIDE SEQUENCE OF 218-257.		
RX		MEDLINE=99043341; PubMed=9827539; DOI=10.1016/S0014-5793(98)01293-9;		
RA		Wagener R., Kobbe B., Paulsson M.;		
RT		"Genomic organisation, alternative splicing and primary structure of		
RT		human matrilin-4."		
RL		FEBS Lett. 438:165-170(1998).		
CC		-I- FUNCTION: Major component of the extracellular matrix of		
CC		cartilage.		
CC		-I- SUBCELLULAR LOCATION: Secreted.		
CC		-I- ALTERNATIVE PRODUCTS:		
CC		Event=Alternative splicing; Named isoforms=2;		
CC		Name=long;		
CC		Isoid=089029-1; Sequence=Displayed;		
CC		Name=short;		
CC		Isoid=089029-2; Sequence=VSP 001401;		
CC		-I- TISSUE SPECIFICITY: Lung, brain, sternum, kidney and heart.		
CC		-I- DEVELOPMENTAL STAGE: The short isoform was detected in 7 weeks old		
CC		mice but not in developing mice (19.5 dpc embryos or in 2, 8, and		
CC		21 days old animals).		
CC		-I- SIMILARITY: Contains 4 EGF-like domains.		
CC		-I- SIMILARITY: Contains 2 VWFA domains.		
CC		-----		
CC		This Swiss-Prot entry is copyright. It is produced through a collaboration		
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC		the European Bioinformatics Institute. There are no restrictions on its		
CC		use as long as its content is in no way modified and this statement is not		

CC	removed.	-----
CC	removed.	-----
DR	EMBL; AJ006140; CAA06889.1; -; mRNA.	
DR	EMBL; AJ006140; CAA06890.1; -; mRNA.	
DR	EMBL; BC036558; AAH36558.1; -; mRNA.	
DR	EMBL; AJ010984; CAA09451.1; -; Genomic_DNA.	
DR	HSSP; P00736; IAP0.	
DR	Ensembl; ENSMUSG00000016995; Mus musculus.	
DR	GO; MGI:1328314; Matn4.	
DR	GO; GO:0005615; C:extracellular space; TAS.	
DR	InterPro; IPR000152; Asx hydroxyl_S.	
DR	InterPro; IPR000742; EGF_2.	
DR	InterPro; IPR006209; EGF_like.	
DR	InterPro; IPR011203; Matn_VA.	
DR	InterPro; IPR02035; VWF_A.	
DR	Pfam; PF00009; EGF_4.	
DR	Pfam; PF00092; VWA; 2.	
DR	PIRSE; PIRSF015217; Matn_VWA; 1.	
DR	PRINTS; PR00453; VWFADOMAIN.	
DR	PROSITE; PS00010; ASX_HYDROXYL; 2.	
DR	PROSITE; PS01186; EGF_2; 2.	
DR	PROSITE; PS50026; EGF_3; FALSE_NEG.	
DR	PROSITE; PS50234; VWF_A; 2.	
KW	Alternative splicing; Coiled coil; EGF-like domain; Glycoprotein; Polymorphism; Repeat; Signal.	
FT	SIGNAL	1 21 Potential.
FT	CHAIN	22 624 Matrilin-4.
FT	DOMAIN	36 215 VWF_A 1.
FT	DOMAIN	217 257 EGF-like 1.
FT	DOMAIN	258 298 EGF-like 2.
FT	DOMAIN	299 339 EGF-like 3.
FT	DOMAIN	340 380 EGF-like 4.
FT	DOMAIN	388 563 VWF_A 2.
FT	COILED	590 623 Potential.
FT	CARBOHYD	71 71 N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	307 307 N-linked (GlcNAc...) (Potential).
FT	DISULFID	221 232 By similarity.
FT	DISULFID	228 241 By similarity.
FT	DISULFID	243 256 By similarity.
FT	DISULFID	262 273 By similarity.
FT	DISULFID	269 282 By similarity.
FT	DISULFID	284 297 By similarity.
FT	DISULFID	303 314 By similarity.
FT	DISULFID	310 323 By similarity.
FT	DISULFID	325 338 By similarity.
FT	DISULFID	344 355 By similarity.
FT	DISULFID	351 364 By similarity.
FT	DISULFID	366 379 By similarity.
FT	VARSPLIC	28 217 Missing (in isoform Short).
FT	VARIANT	319 319 A -> E (in strain C57BL/6J).
FT	VARIANT	346 346 G -> D (in strain C57BL/6J).
SEQ	SEQUENCE	624 AA; 68918 MW; DPA28D2C94B1A14F CRG64;
Alignment Scores:		
Pred. No.:	1.5e-11	Length: 624
Score:	265.00	Matches: 64
Percent Similarity:	53.4%	Conservative: 29
Best Local Similarity:	36.8%	Mismatches: 75
Query Match:	25.3%	Indels: 6
DB:	1	Gaps: 2
US-10-699-035A-1 (1-537) x MATN4_MOUSE (1-624)		
QY	4 GACCTGATGTTCTCTGCTGTGACAGCTTCACGACGATTCACATGAGATTCTCCGGATT	63
DB	36 AspLeuValPheMetIleAspSerSerValArgProPheGluPheGluThrMet	55
QY	64 CGGAGTTGTTGGGGGAGCTGGGCTTCACGACGCTCGGGACCGGGGCGCTGGCC	123
DB	56 ArgGlnPheLeuValGlyIleuLeuArgSerLeuAspValGlyIleuAsnAlaThrArgVal	75
QY	124 AGTCGTGTGACGCTGGGGAGCTGGCCATACCGAGTTCCTTCGAGCAGACAGACTCG	183


```

Db      76 GlyValIleGlnTyrSerSerGlnValGlnSerValPheProLeuGlyAlaPheSerArg 95
Oy      184 GGTGAGGCTGCCAGATGCGGTGGCTGCTTGTGCCAGCCGATGGGTGACACCACT 243
Db      96 ArgGluAspMetGlnArgAlaIleArgAlaValValProLeuAlaGlnGlyThrMetThr 115
Oy      244 GGCCTGGCGCTGCTGATGCCAAGACAGCTGTTGCTGAAGCATCAAGTCCCGGCCA 303
Db      116 GlyLeuAlaIleGlnTyrAlaMetAsnValAlaPheSerGlnAlaGlnGlyAlaArgPro 135
Oy      304 GGG-----GTGCCCAAAAGTCTGCTGTGGGTGACGATGGCGGCTCCAGCACT 354
Db      136 SerGluGluArgValProArgValLeuValIleValThrAspGlyArgProGlnAspArg 155
Oy      355 GTGGGCCCCCGCATGACAGAGCTCAAGAGCACTGGCGGCTGTCATGTCACAC 414
Db      156 ValAlaGluValAlaAlaGlnAlaArgAlaArgGlyIleGluIleTyrAlaValGlyVal 175
Oy      415 GGCCTGAGGCACTTCTGAGCTGTACGCCGCTGCTCAAGCCCTGCCAGAGACCTGT 474
Db      176 GlnArgAlaAspValGlySerLeuArgThrMetAlaSerProProLeuAspGlnHisVal 195
Oy      475 CACTTGTGACGTGATGACCTGCACATCATTTGTCAAAG 516
Db      196 PheLeuValGluSerPheAspLeu-----IleGlnGlu 206

RESULT 12
COCAL CHICK STANDARD; PRT; 3124 AA.
ID      COCAL CHICK 013944; 004509;
AC      01-JAN-1990 (Rel. 13, Created)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      10-MAY-2005 (Rel. 47, Last annotation update)
GN      Collagen alpha 1(XII) chain precursor (Fibrochimerin).
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianae;
OC      Gallus.
CX      NCBI_TaxID=9031;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=White leghorn;
RX      MEDLINE=92011862; PubMed=1918137; DOI=10.1083/jcb.115.1.209;
RA      Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA      Nishida Y., Obata M., Kimata K.;
RT      "The complete primary structure of type XII collagen shows a chimeric
RT      molecule with reiterated fibronectin type III motifs, von Willebrand
RT      factor A motifs, a domain homologous to a noncollagenous region of
RT      type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT      site."
RT      J. Cell Biol. 115:209-221 (1991).
RN      [2]
RP      NUCLEOTIDE SEQUENCE OF 2456-3124, AND PROTEIN SEQUENCE OF 2772-2794
RP      AND 2846-2873.
RX      MEDLINE=90062079; PubMed=2584192;
RA      Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
RT      "Type XII collagen. A large multidomain molecule with partial homology
RT      to type IX collagen."
RT      J. Biol. Chem. 264:19772-19778 (1989).
RN      [3]
RP      NUCLEOTIDE SEQUENCE OF 2960-3076.
RX      MEDLINE=87317590; PubMed=3476925;
RA      Gordon M.K., Gerecke D.R., Olsen B.R.;
RT      "Type XII collagen: distinct extracellular matrix component discovered
RT      by cDNA cloning."
RT      Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044 (1987).
RN      [4]
RP      NUCLEOTIDE SEQUENCE OF 1-1283 (ISOFORM SHORT), AND ALTERNATIVE
RP      SPLICING.
RC      TISSUE=Embryo;
RX      MEDLINE=93042014; PubMed=1420368; DOI=10.1016/0167-4781(92)90145-P;

```

```

RA      Trueb J., Trueb B.;
RT      "The two splice variants of collagen XII share a common 5' end.";
RT      Biochim. Biophys. Acta 1171:97-98 (1992).
RN      [5].
RP      ALTERNATIVE SPLICING.
RX      MEDLINE=95370352; PubMed=7642694; DOI=10.1083/jcb.130.4.1005;
RA      Koch M., Bohmann B., Mathison M., Hagios C., Trueb B., Chiquet M.;
RT      "Large and small splice variants of collagen XII: differential
RT      expression and ligand binding."
RT      J. Cell Biol. 130:1005-1014 (1995).
CC      -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC      containing fibrils, the COL1 domain could be associated with the
CC      surface of the fibrils, and the COL2 and NC3 domains may be
CC      localized in the perifibrillar matrix.
CC      -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC      nontriple-helical sequences.
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Comment=The final tissue form of collagen XII may contain
CC      homotrimers of either isoform long or isoform short or any
CC      combination of isoform long and isoform short. Only isoform long
CC      is a proteoglycan. Isoform long has more restricted expression
CC      in embryonic tissue than isoform short;
CC      Name=Long;
CC      IsoId=PI3944-1; Sequence=Displayed;
CC      Name=Short;
CC      IsoId=PI3944-2; Sequence=VSP 001148;
CC      -1- TISSUE SPECIFICITY: Type XII collagen is present in tendons,
CC      ligaments, perichondrium, and perosteum, all dense connective
CC      tissues containing type I collagen.
CC      -1- DOMAIN: This sequence defines five distinct domains, two triple-
CC      helical domains (COL1 and COL2) and three nontriple-helical
CC      domains (NC1, NC2, and NC3).
CC      -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC      each end.
CC      -1- PTM: Prolines at the third position of the tripeptide repeating
CC      unit (G-X-Y) are hydroxylated in some or all of the chains.
CC      -1- PTM: O-glycosylated; glycosaminoglycan of chondroitin-sulfate type
CC      (By similarity).
CC      -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC      interrupted helices (FACIT) family.
CC      -1- SIMILARITY: Contains 18 fibronectin type-III domains.
CC      -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC      -1- SIMILARITY: Contains 4 WFPA domains.
CC      -----
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use as long as its content is in no way modified and this statement is not
CC      removed.
CC      -----
DR      EMBL, D00824; BAA00701.1; -; mRNA.
DR      EMBL, X61024; CAA43358.1; -; mRNA.
DR      EMBL, M17375; AAA48718.1; -; mRNA.
DR      EMBL, J05137; AAA48635.1; -; mRNA.
DR      EMBL, X67327; CAA47744.1; -; mRNA.
DR      PIR, A40020; A40020.
DR      HSSP, P56199; 10C5.
DR      Ensemble; ENSGALG00000015908; Gallus gallus.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR003961; FN_III.
DR      InterPro; IPR003129; Laminin_G_TSP_N.
DR      InterPro; IPR002035; WVF_A.
DR      Pfam, PF01391; Collagen; 4.
DR      Pfam, PF00041; Fn3; 17.
DR      Pfam, PF00092; WVA; 4.
DR      PRINTS, PR00453; WVFADOMAIN.
DR      SMART, SMO0060; FN3; 18.
DR      SMART, SMO0210; TSPN; 1.
DR      SMART, SMO0327; WVA; 4.
DR      PROSITE, PS50853; FN3; 18.
DR      PROSITE, PS50234; WVFA; 4.
KW      Alternative splicing; Cell adhesion; Collagen;

```

KW Direct protein sequencing; Extracellular matrix; Glycoprotein;
 KM Hydroxylation; Repeat; Signal; Structural protein.
 FT SIGNAL 1 24 Potential.
 FT CHAIN 25 3124 Collagen alpha 1(XII) chain.
 FT DOMAIN 25 112 Fibronectin type-III 1.
 FT DOMAIN 139 311 VFMA 1.
 FT DOMAIN 332 421 Fibronectin type-III 2.
 FT DOMAIN 439 615 VFMA 2.
 FT DOMAIN 630 718 Fibronectin type-III 3.
 FT DOMAIN 721 809 Fibronectin type-III 4.
 FT DOMAIN 812 902 Fibronectin type-III 5.
 FT DOMAIN 905 993 Fibronectin type-III 6.
 FT DOMAIN 995 1083 Fibronectin type-III 7.
 FT DOMAIN 1086 1175 Fibronectin type-III 8.
 FT DOMAIN 1199 1371 VFMA 3.
 FT DOMAIN 1386 1472 Fibronectin type-III 9.
 FT DOMAIN 1474 1564 Fibronectin type-III 10.
 FT DOMAIN 1566 1654 Fibronectin type-III 11.
 FT DOMAIN 1655 1745 Fibronectin type-III 12.
 FT DOMAIN 1756 1845 Fibronectin type-III 13.
 FT DOMAIN 1847 1935 Fibronectin type-III 14.
 FT DOMAIN 1937 2026 Fibronectin type-III 15.
 FT DOMAIN 2028 2117 Fibronectin type-III 16.
 FT DOMAIN 2119 2206 Fibronectin type-III 17.
 FT DOMAIN 2210 2294 Fibronectin type-III 18.
 FT DOMAIN 2324 2500 TSP N-terminal.
 FT DOMAIN 2524 2716 Nonhelical region (NC3).
 FT REGION 2751 2902 Triple-helical region (COL2) with 1
 FT REGION 2903 2945 Triple-helical region (NC2).
 FT REGION 2946 3048 Nonhelical region (COL1) with 2
 FT REGION 3049 3124 Imperfections.
 FT MOTIF 2899 2901 Nonhelical region (NC1).
 FT COMPBIAS 3086 3096 Cell attachment site (Potential).
 FT COMPBIAS 3111 3123 Arg/Lys-rich (acidic).
 FT CARBOHYD 32 32 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 797 797 O-linked (Xyl. . .) (chondroitin sulfate) (Potential).
 FT CARBOHYD 890 890 O-linked (Xyl. . .) (chondroitin sulfate) (Potential).
 FT CARBOHYD 981 981 O-linked (Xyl. . .) (chondroitin sulfate) (Potential).
 FT CARBOHYD 1006 1006 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1032 1032 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1044 1044 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1512 1512 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1767 1767 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 2210 2210 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 2273 2273 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 2532 2532 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 2683 2683 N-linked (GlcNAc. . .) (Potential).
 FT VARSPLIC 25 1188 Missing (in isoform short).
 FT CONFLICT 1258 1258 /FTid=VSP_001148.
 FT CONFLICT 1264 1264 T -> S (in Ref. 4).
 FT CONFLICT 2759 2759 D -> E (in Ref. 4).
 FT CONFLICT 2803 2803 P -> A (in Ref. 2).
 FT CONFLICT 2977 2977 L -> F (in Ref. 2).
 FT CONFLICT 2977 2977 V -> F (in Ref. 2).
 FT CONFLICT 3075 3076 QP -> AG (in Ref. 3).
 SQ SEQUENCE 3124 AA; 340582 MW; 094285AF67E346CF CRC64;

Alignment Scores:
 Pred. No.: 2,23e-11 Length: 3124
 Score: 263.50 Matches: 62
 Percent Similarity: 53.1% Conservative: 32
 Best Local Similarity: 35.0% Mismatches: 82
 Query Match: 25.1% Indels: 1
 DB: 1 Gaps: 1

US-10-699-035a-1 (1-537) x COCA1_CHICK (1-3124)

OY 4 GACCTGATGTTCCCTGCTGAGACGTCAGCCAGCCTCTCTACATACAGATTCTCCCGGTT 63
 DB |||||
 OY 439 ASPVAlValPheLeuValAspGlySerTyrSerIleGlyLeuAlaAspPheValValVal 458
 DB |||||
 OY 64 CGGAGGTTGGGGGCGAGCTGGTGGCTCCAGCTGCCCGCCGAGCGGAGCCGCTGATGCC 123
 DB |||||
 DB 459 ArgAlaPheLeuGluValLeuValLysSerPheGluLysSerProAlaGlyValGlnLe 478
 OY 124 AGTGTGGTGCAGCTGGGCGAGTGGCGCATACCCAGATGTTCCCTTCGGCGAGCAGATCG 183
 DB |||||
 DB 479 SerLeuValGlnTyrSerArgAspProIleMetGluPheSerLeuAlaArgTyrAsnArg 498
 OY 184 GGTAGGCTCCCGAGATCCGCTGCTTCTTCCCGAGCGATGGTGAACCCACACT 243
 DB ::|||
 DB 499 ValLysAspIleIleGlnAlaIleAsnThrPheProTyrArgGlyGlySerThrAsnThr 518
 OY 244 GGCCTGGCGCTGCTGTATGCAAGAAAGCTGTTGTCGATGATCAGTGGCCGCGCA 303
 DB |||||
 DB 519 GlyLysAlaMetThrTyrValArgLysValPheValThrSerLysGlySerArgPro 538
 OY 304 GGGGTGCCAAAGTGTGCTGGTGGTGAACAGATGGCGCTCCAGCAGCCTGTGGGCCCC 363
 DB |||||
 DB 539 AsnValProArgValMetIleLeuIleThrAspGlyLysSerSerAspAlaPheLysGlu 558
 OY 364 CCATGACGAGCTCAAGAGACCTGGCGCTCACCGTTCATTGTCAGACCGCGCGAGGC 423
 DB ::|||
 DB 559 ProAlaIleLysLeuArgAspAlaAspValGluIlePheAlaValGlyValLysAspAla 578
 OY 424 AACCTTCCTGAGCTGTACCGCTGCTGCTGACCGCTGCGAGGAGACCTGACCTTGTG 483
 DB |||||
 DB 579 ValArgThrGluLeuGluValIleAlaIleSerProAlaGluThrAlaValTyrThrVal 598
 OY 484 --GACGTGATGACCTGACATCATGTTCCAGAGCTGAGGCGCTCAT 531
 DB |||||
 DB 599 GluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerVal 615

RESULT 13
 Q4SH63_TESTNG PRELIMINARY; PRT; 1557 AA.
 ID Q4SH63_TESTNG
 AC Q4SH63;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DE Chromosome 8 SCAFL14587, whole genome shotgun sequence.
 DE (Fragment).
 DE ORFNames=GSTENG0018311001;
 OS Tetraodon nigroviridis (Green puffer).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicoud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dastivla C., Salanoubat M., Levy M., Boudet N., Castellan S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Catroilo L., Poulain J., de Bernardis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosaek S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Landel V., Schachter V., Queller F., Saurin W., Scarpelli C.,
 RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.,
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (Feb-2004) to the EMBL/Genbank/DBJ databases.

CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CA601014587; CAG00019.1; -; Genomic_DNA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00092; VMA; 2.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VMA; 2.
DR PROSITE; PS50853; FN3; 7.
DR PROSITE; PS50234; VWF_A; 2.
KM Collagen; Extracellular matrix; Structural protein.
FT NON_TER
SQ SEQUENCE 1557 AA; 168266 MW; 6203E9A0900C6D9D CRC64;

Alignment Scores:
Pred. No.: 2,29e-11 Length: 1557
Score: 263.00 Matches: 61
Percent Similarity: 56.3% Conservative: 37
Best Local Similarity: 35.1% Mismatches: 74
Query Match: 25.1% Indels: 2
DB: Gaps: 2

US-10-699-035a-1 (1-537) x Q4SH63_TETNG (1-1557)

QY 4 GACCTGATGTCCTGCTGGACAGCTCAGCGCTCTCACTACGAGTTCTCCGGGTT 63
DB 891 AsplevValPheleuValaspGlyserTpsrIlleGlyAspIuasnheMetLysile 910
QY 64 CGGAGTTGTGGGGGAGCTGTGCTCCACTGCCCTG--GCAACCGGGGCTTCGCT 120
DB 911 ThrArgPheLeuHisSerThrValGlySerLeuAspLeuIleGlyThrAspGlyThrGln 930
QY 121 GCGAGTCGTGGACCTGGGAGCTGGCCATACCGAGTTCCTCCCTGGCGGACGACG 180
DB 931 ValAlaIleAlaGlnPheSerAspAspAlaArgThrGlnPheGlnLeuSerSerHisSer 950
QY 181 TCGGCTGAGGCTCCCGCAGATGCGCTGCTTCTGCCGACGATGGGTGACACCCAC 240
DB 951 AsnLysGlnAlaLeuLeuGlnAlaIleGlnLysIleSerTrpLysGlyLysThrLys 970
QY 241 ACTGCGCTGCGCTGTCTATGCAAGAAACAGCTTTGCTGAAGCATCAGGTGCGCG 300
DB 971 ThrGlyArgAlaIleLysHisValLysGlnSerIlePheSerLeuGlnIleGlyAlaArg 990
QY 301 CCAAGGCTGCCCAAAATGCTGTGTGGTGACGATGGCGGCTCCGACGACCTGTGGC 360
DB 991 ArgGlyValProLysValLeuValValLeuHisAspGlyArgSerGlnAspValAsn 1010
QY 361 CCCCCCATCAGAGCTCAAGACCTGGCGCTGCTTCACTTGTTCAGACCGGCGCA 420
DB 1011 LysValSerLysGlnMetGlnMetAspGlyTrpIleIlePheAlaIleGlyPheAlaAsp 1030
QY 421 GGCACCTTCCTGAGCTGTCAAGCCGCTGCTTACGCCCTGGCGAAGACCTGCACTT 480
DB 1031 AlaAspTrpGlyGlnLeuValAsnIleAlaSerLysProSerAspArgHisValPhePhe 1050
QY 481 GTC--GACGTGATGACCTGCACATCATCTGTCACAGACTG 519
DB 1051 ValAspAspLeuAspAlaValLysLysIleGlnGlnLeu 1064

RESULT 14
OSVYK2_HUMAN PRELIMINARY; PRT; 2884 AA.
AC OSVYK2;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
BT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Collagen, type XII, alpha 1.
GN Name=COL12A1; ORFNames=RP1-238D15.1-003;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OC NCBI_TaxID=9606;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RA Tromans A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN (2)
RP NUCLEOTIDE SEQUENCE.
RA Smith M.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN (3)
RP NUCLEOTIDE SEQUENCE.
RA Cordy N.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL354664; CAH71309.1; -; Genomic_DNA.
DR EMBL; AL080250; CAI19897.1; -; Genomic_DNA.
DR EMBL; AL096771; CAI19907.1; -; Genomic_DNA.
DR EMBL; AL080250; CAH71309.1; JOINED; Genomic_DNA.
DR EMBL; AL354664; CAH71309.1; JOINED; Genomic_DNA.
DR EMBL; AL080250; CAI19897.1; JOINED; Genomic_DNA.
DR EMBL; AL080250; CAI19907.1; JOINED; Genomic_DNA.
DR EMBL; AL354664; CAI19907.1; JOINED; Genomic_DNA.
DR EMBL; AL096771; CAI19897.1; JOINED; Genomic_DNA.
DR EMBL; ENSG00000111799; Homo sapiens.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00041; fn3; 18.
DR Pfam; PF00092; VMA; 4.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00060; FN3; 18.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VMA; 4.
DR PROSITE; PS50853; FN3; 18.
DR PROSITE; PS50234; VWF_A; 4.
KM Collagen; Extracellular matrix; Repeat; Structural protein.
SQ SEQUENCE 2884 AA; 315869 MW; 2D598F1365EB454D CRC64;

Alignment Scores:
Pred. No.: 3.14e-11 Length: 2884
Score: 261.50 Matches: 64
Percent Similarity: 52.5% Conservative: 29
Best Local Similarity: 36.2% Mismatches: 83
Query Match: 24.9% Indels: 1
DB: Gaps: 1

US-10-699-035a-1 (1-537) x OSVYK2_HUMAN (1-2884)

QY 4 GACCTGATGTCCTGCTGGACAGCTCAGCGCTCTCACTACGAGTTCTCCGGGTT 63
DB 440 AsplevValPheleuValaspGlyserTrpIleGlyIleAlaIleValLysVal 459
QY 64 CGGAGTTGTGGGGGAGCTGTGCTCCACTGCCCTGGGACGCGGGGCTTCGCTGCC 123
DB 460 ArgAlaPheLeuGlnValLeuValLysSerPheGlnIleSerProAsnArgValGlnIle 479
QY 124 ACTGTGTCGACAGTGGGACGCTGGCGCATCAGCGAGTTCCCTTGGCGGACAGACTG 183
DB 480 SerLeuValGlnTrpSerArgAspProHisTrpGlnPheThrLeuLysPheThrLys 499

```

QY 184 GGTAGGCTCCAGATCGCGTGGCTTCTGCCCCAGCCATGGTGACACCACACT 243
D 500 ValGluSplrIleIleGluAlaIleAsnThrPheProTyrArgGlySerThrAsnThr 519
QY 244 GGCGTGGCGGTGTGTCCAGAGACAGCGTTGTCTGTAAGATCAGTGGCCCGGCA 303
D 520 GlySalMetCmTyrValArgGluValIlePheValProSerLysGlySerGser 539
QY 304 GGGGTGCCCAAGTCTGTGTGTGACAGATGGCGGCTCCAGGACCTGTGGGCCCC 363
D 540 AsnValProLysValMetIleLeuIleThrAspGlyLysSerSerAspAlaPheArgAsp 559
QY 364 CCCATGACGAGGCTCAGACAGCTGGGCGTCAACCGTTTCATTGTGACAGCCGCGAGGC 423
D 560 ProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValLysAspAla 579
QY 424 AACTTCCTGGAGCTGTGACCGCTCCAGCCCTGCGGAGAGACCTGCACCTTTGTG 483
D 580 ValArgSerGluLeuGluAlaIleAlaSerProProAlaGluThrHisValPheThrVal 599
QY 484 ---GACGTGATGACCTGCACATCATTTGTCCAGAGAGTGGGCGCTCCATT 531
D 600 GluAspPheAspAlaPheGluArgIleSerPheGluLeuThrGlnSerIle 616

RESULT 15
COCAL HUMAN STANDARD; PRT; 3063 AA.
ID COCAL HUMAN 099715; 099716;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN Name=COL12A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=97286521; PubMed=9143499; DOI=10.1006/geno.1997.4638;
RA Gerecke D.R., Olson P.F., Koch W., Knoll J.H.M., Taylor R.,
RA Hudson D.L., Champilaud M.-F., Olsen B.R., Burgesson R.E.;
RT "Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(XI)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL9A1) to human
RT chromosome 6q12-q13."
RL Genomics 41:236-242(1997).
CC -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the pericellular matrix (By similarity).
CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=The final tissue form of collagen XII may contain
CC homotrimers of either isoform long or isoform short or any
CC combination of isoform long and isoform short;
CC Name=Long;
CC IsoId=Q99715-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q99715-2; Sequence=VSP_001149;
CC -1- TISSUE SPECIFICITY: Found in collagen I-containing tissues: both
CC isoform short and isoform long appear in amnion, chorion, skeletal
CC muscle, small intestine, and in cell culture of dermal
CC fibroblasts, keratinocytes and endothelial cells. Only the short
CC isoform is found in lung, placenta, kidney and a squamous cell
CC carcinoma cell line.
CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end (By similarity).
CC -1- PTM: Prolines at the third position of the tripeptide repeating

```

```

CC unit (G-X-Y) are hydroxylated in some or all of the chains (By
CC similarity).
CC -1- PTM: O-glycosylation of isoform long; glycosaminoglycan of
CC chondroitin-sulfate type (By similarity).
CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC interrupted helices (FACIT) family.
CC -1- SIMILARITY: Contains 18 fibronectin type-III domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -1- SIMILARITY: Contains 4 VWFA domains.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL: U73778; AAC51244.1; -; mRNA.
DR EMBL: U73779; AAD40483.1; -; mRNA.
DR HSSP: P18614; IMHP.
DR Ensembl: ENSG00000111799; Homo sapiens.
DR HGN: HGNC:2188; COL12A1.
DR MIM: 120320; -.
DR GO: GO:0005595; C:collagen type XII; TAS.
DR GO: GO:0001501; P:skeletal development; TAS.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR002035; VWF A.
DR Pfam: PF01391; Collagen; 4.
DR Pfam: PF00041; fn3; 18.
DR Pfam: PF00092; VWF; 4.
DR PRINTS: PR00453; VWFADOMAIN.
DR PROSITE: PS50853; FN3; 18.
DR PROSITE: PS50234; VWFA; 4.
DR Alternative splicing; Cell
DR Direct protein sequencing; Extracellular matrix; Glycoprotein;
KW Hydroxylation; Repeat; Signal; Structural protein.
FT SIGNAL 1 24
FT CHAIN 1 24
FT DOMAIN 25 3063
FT DOMAIN 25 112
FT DOMAIN 140 316
FT DOMAIN 333 422
FT DOMAIN 440 616
FT DOMAIN 631 719
FT DOMAIN 722 810
FT DOMAIN 813 901
FT DOMAIN 904 993
FT DOMAIN 995 1083
FT DOMAIN 1086 1175
FT DOMAIN 1199 1371
FT DOMAIN 1384 1472
FT DOMAIN 1474 1563
FT DOMAIN 1565 1652
FT DOMAIN 1654 1743
FT DOMAIN 1752 1841
FT DOMAIN 1843 1931
FT DOMAIN 1933 2022
FT DOMAIN 2024 2113
FT DOMAIN 2115 2202
FT DOMAIN 2206 2290
FT DOMAIN 2323 2496
FT DOMAIN 2520 2712
FT DOMAIN 2745 2746
FT REGION 2747 2898
FT REGION 2899 2941
FT REGION 2942 3044
FT REGION 3045 3063
FT MOTIF 862 864
FT MOTIF 2779 2781
FT MOTIF 2895 2897
FT MOD_RES 2944 2944
FT MOD_RES 2947 2947
CC -----
CC TSP N-terminal.
CC Nonhelical region (NC3).
CC Triple-helical region (COL2) with 1
CC imperfection.
CC Nonhelical region (NC2).
CC Triple-helical region (COL1) with 2
CC imperfections.
CC Nonhelical region (NC1).
CC Cell attachment site (Potential).
CC Cell attachment site (Potential).
CC Cell attachment site (Potential).
CC Hydroxyproline (By similarity).
CC Hydroxyproline (By similarity).

```

Search completed: February 13, 2006, 13:38:53
Job time : 147.574 secs

FT MOD_RES 2950 2950 Hydroxyproline (By similarity).
FT MOD_RES 2959 2959 Hydroxyproline (By similarity).
FT MOD_RES 2965 2965 Hydroxyproline (By similarity).
FT MOD_RES 2968 2968 Hydroxyproline (By similarity).
FT MOD_RES 2971 2971 Hydroxyproline (By similarity).
FT MOD_RES 2983 2983 Hydroxyproline (By similarity).
FT MOD_RES 3000 3000 Hydroxyproline (By similarity).
FT MOD_RES 3003 3003 Hydroxyproline (By similarity).
FT MOD_RES 3014 3014 Hydroxyproline (By similarity).
FT MOD_RES 3023 3023 Hydroxyproline (By similarity).
FT MOD_RES 3026 3026 Hydroxyproline (By similarity).
FT MOD_RES 3029 3029 Hydroxyproline (By similarity).
FT CARBOHYD 700 700 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 798 798 O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT CARBOHYD 889 889 O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT CARBOHYD 981 981 O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT CARBOHYD 1763 1763 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2206 2206 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2528 2528 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2679 2679 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 25 1188 Missing (in isoform Short).
SQ SEQUENCE 3063 AA; 333194 MW; 75FEA78FA8E48293 CRC64;

Alignment Scores:
Pred. No.: 3,16e-11 Length: 3063
Score: 261.50 Matches: 64
Percent Similarity: 52.5% Conservative: 29
Best Local Similarity: 36.2% Mismatches: 83
Query Match: 24.9% Indels: 1
DB: 1 Gaps: 1

US-10-699-035A-1 (1-537) x COCAI_HUMAN (1-3063)

QY 4 GACCTGATGTTCTGCTGAGACAGAGCTCTCACTACGAGTTCTCCGGATT 63
DB 440 AAPIIleValPheLeuValAspGlySerTyrSerIleGlyIleAlaSerPheValVal 459
QY 64 CGGGAGTTTGTGGGGCAGCTGCTGCTCACTGCCCTGGGACCGGGGCTGCTGCC 123
DB 460 ArgAlaPheLeuGluValLeuValLysSerPheGluIleSerProAsnArgValGluIle 479
QY 124 ACTCTGTCGACAGTGGGACAGTGGCCATACACCGAGTTCCCTTGGCCAGCAGACTCG 183
DB 480 SerLeuValGlnTyrSerArgAspProHisThrGluPheThrLeuLysPheThrLys 499
QY 184 GGTGAGGCTGCCGAGATGGGGTGGCTTTCGCCGAGGATGGGTGACACCCGACT 243
DB 500 ValGluAspIleIleGluAlaIleAsnThrPheProTyrArgGlySerThrAsnThr 519
QY 244 GGCCTGGCCCTGCTATGCAAGAACAGCTGTTTCTGAAGCATCAGGTGCCGAGCA 303
DB 520 GlyLysAlaMetThrTyrValArgGluLysIlePheValProSerLysGlySerArgSer 539
QY 304 GGGGTGCCCAAGTGTGCTGAGTGGTGAAGATGGCGGCTCCAGCAGCCCTGGGGCCC 363
DB 540 AsnValProLysValMetIleLeuIleThrAspGlyLysSerSerAspAlaPheArgAsp 559
QY 364 CCCATGCAAGAGCTCAAGAGCCTGGGGCTGCTTCATGTCAGACACCGGCGAGGC 423
DB 560 ProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValLysAspAla 579
QY 424 AACTTCTGAGAGCTGTCAAGCCGTGCTCAAGCCCTGCGAGAGCAGCTGCACTTTGTG 483
DB 580 ValArgSerGluLeuGluAlaIleAlaSerProProAlaGluThrHisValPheThrVal 599
QY 484 ---GACGTGATGACCTGACATCATTTGCAAGAGCTGAGGGGCTTCATT 531
DB 600 GluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIle 616

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:12:53 ; Search time 18.5896 Seconds
(without alignments)
2538.475 Million cell updates/sec

Title: US-10-699-035A-1

Perfect score: 1049
Sequence: 1 999gaccctgctctctcgcctcctcctcctcctcctcgcg 537

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2443163 seqs, 439378781 residues
Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-O=/abs/ABSSWB spool/US10699035/runat_13022006_062440_25403/app_query.fasta_1
-DB=A.Geneseq -GFM=factan -SUFFIX=rag -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.coi -LIST=45
-DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=US10699035@CGN 1.1.476 @runat_13022006_062440_25403 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	902	86.0	180	6	AAE32500	AAE32500 Human von
2	902	86.0	215	4	AAE87344	AAE87344 Human gen
3	902	86.0	215	5	ABG65347	ABG65347 Human alb
4	902	86.0	215	8	ADL78614	ADL78614 Albunin f
5	902	86.0	226	4	AAE87424	AAE87424 Human gen
6	902	86.0	242	4	AAE87418	AAE87418 Human gen
7	902	86.0	285	9	AAE88340	AAE88340 Human mem
8	902	86.0	285	9	ADY63045	ADY63045 Human c10
9	902	86.0	418	6	AAE32502	AAE32502 Human Wll

10	902	86.0	445	5	ABE69674	ABE69674 Human pol
11	902	86.0	445	8	ADH71106	ADH71106 Human pro
12	886	84.5	185	4	AAE03654	AAE03654 Human ext
13	877	83.6	299	3	AAE42581	AAE42581 Human ORF
14	762	72.6	180	6	AAE32503	AAE32503 Mouse von
15	762	72.6	415	6	AAE32501	AAE32501 Mouse Wll
16	754	71.9	421	2	AAE86326	AAE86326 Kidney in
17	751	71.6	186	4	AAE87419	AAE87419 Human gen
18	384	36.6	176	3	AAE02196	AAE02196 Human sec
19	272.5	26.0	176	5	AAE78816	AAE78816 Von Wille
20	265	25.3	644	8	ADY66963	ADY66963 Mouse can
21	265	25.3	644	8	ADY12632	ADY12632 Murine can
22	261.5	24.9	3063	5	ABE90762	ABE90762 Human Tum
23	261.5	24.9	3063	6	ABE05469	ABE05469 Human tum
24	261.5	24.9	3063	6	ABE47415	ABE47415 Breast ca
25	261.5	24.9	3063	6	ABE47416	ABE47416 Breast ca
26	261.5	24.9	3063	6	ADY75666	ADY75666 Marker ge
27	261.5	24.9	3118	4	AAU27790	AAU27790 Human ful
28	259	24.7	293	3	AAE39142	AAE39142 Human sec
29	259	24.7	357	3	AAE93189	AAE93189 Human pro
30	259	24.7	357	8	ADL30530	ADL30530 Human pro
31	259	24.7	391	9	ADY12639	ADY12639 Human can
32	259	24.7	488	4	AAE93707	AAE93707 Human pol
33	259	24.7	488	8	ADL31608	ADL31608 Human pro
34	259	24.7	499	9	ADY12635	ADY12635 Human can
35	259	24.7	499	9	ADY12643	ADY12643 Human can
36	259	24.7	540	9	ADY12637	ADY12637 Human can
37	259	24.7	540	9	ADY12641	ADY12641 Human can
38	259	24.7	580	8	ADY19949	ADY19949 Human sof
39	259	24.7	581	8	ADY66966	ADY66966 Human can
40	259	24.7	581	9	ADY12645	ADY12645 Human can
41	259	24.7	581	9	ADY12647	ADY12647 Human can
42	259	24.7	620	9	ADY12649	ADY12649 Human can
43	254.5	24.3	795	5	AAU84267	AAU84267 Human end
44	254.5	24.3	795	4	AAE27229	AAE27229 Human EXM
45	254.5	24.3	1207	7	ADY69785	ADY69785 Human hea

ALIGNMENTS

RESULT 1	AAE32500	standard; protein; 180 AA.
ID	AAE32500	standard; protein; 180 AA.
XX	AAE32500;	
AC	AAE32500;	
XX	AAE32500;	
DT	24-MAR-2003	(first entry)
XX	24-MAR-2003	(first entry)
DE	Human von Willebrand Factor A (VA) domain.	
XX	Human von Willebrand Factor A (VA) domain.	
KW	Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;	
KW	extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;	
XX	gene therapy; human.	
OS	Homo sapiens.	
XX	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	Misc-difference 1..2	
FT	Misc-difference 179..180	/note="Encoded by GGS"
FT	Misc-difference 179..180	/note="Encoded by CTC"
PN	WO200288184-A1.	
XX	WO200288184-A1.	
PD	07-NOV-2002.	
XX	07-NOV-2002.	
PF	02-MAY-2002; 2002WO-AU000542.	
XX	02-MAY-2002; 2002WO-AU000542.	
PR	02-MAY-2001; 2001AU-00004701.	
XX	02-MAY-2001; 2001AU-00004701.	
PA	(MURD-) MURDOCH CHILDRENS RES INST.	
XX	(MURD-) MURDOCH CHILDRENS RES INST.	
PI	Bateman JF, Fitzgerald DJ;	

XX AC WPI: 2003-111873/10.
 XX DR N-PSDB; AAD50397.
 XX PT New isolated Willebrand Factor A-related protein polypeptide useful for
 XX the manufacture of a medicament in the treatment of a disease condition
 XX of the extracellular matrix, in particular arthritis.
 XX PS Claim 7, Page 72-73; 103pp; English.
 XX CC The invention relates to Willebrand Factor A domain related-protein
 XX (WARP) which is a member of von Willebrand factor A (vA)-domain protein
 XX superfamily of extracellular matrix (ECM) proteins. WARP is used as a
 XX molecular marker, used for detecting a loss of ECM integrity in an animal
 XX subject, monitoring repair, regeneration or other disease processes in an
 XX animal subject and detecting a disease condition or a propensity for the
 XX development of a disease condition in an animal subject. The invention is
 XX useful for the manufacture of a medicament in the treatment of a disease
 XX condition of the ECM. The disease condition involves the cartilage, and
 XX is preferably arthritis. The invention is also used in gene therapy. The
 XX present sequence is human vA domain
 XX SQ Sequence 180 AA;
 Alignment Scores:
 Pred. No.: 4,51e-69 Length: 180
 Score: 902.00 Matches: 178
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 6 Gaps: 0
 US-10-699-035A-1 (1-537) x AAB82500 (1-180)
 QY 1 GGGGACCTGATGTTCTCTGTCGACAGCTACGCCGCTCTCTCACTACAGATTCTCCCGG 60
 DB 2 G1YAspLeuMetPheLeuLeuAAspSerSerAlaSerValSerHisTyrGluPheSerArg 21
 QY 61 GTTCGGAGATTGTCGGGAGCTGTCGCTGCTCACTGCCCCGAGGACCGGAGCCCTGCGT 120
 DB 22 ValArgGluPheValG1YglnLeuValAlaProLeuProLeuG1YThrG1YAlaLeuArg 41
 QY 121 GCCAGTCTGTGTCAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 180
 DB 42 AlaSerLeuValHisValG1YSerArgProTyrThrGluPheProPheG1YAlaHisSer 61
 QY 181 TCGGGTGAAGCTGCCCGACGATGCGGTCGCTCTCTGCCCCGACGATGCGTGACACCCAC 240
 DB 62 SerG1YglnAlaAlaGlnAAspAlaValArgAlaSerAlaGlnArgMetG1YAspThrHis 81
 QY 241 ACTGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 300
 DB 82 ThrG1YLeuAlaLeuValTyrAlaLeuSerG1YglnLeuPheAlaG1YAlaHisSerG1YAlaArg 101
 QY 301 CCAGGGGTGCCCAAGTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 360
 DB 102 ProG1YValProLysValLeuValTyrValThrAspG1YglnSerSerAspProValG1Y 121
 QY 361 CCCCCATGTCAGAGACTCAAGACCTGGGCGTCACCCGTCCTCATTTGTCAGACCGGCGCA 420
 DB 142 G1YAspPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 161
 QY 421 GGCAACTTCCTGGAGCTGTCAGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 480
 DB 162 ValAspValAspAspLeuHisIleLeuValGlnGlnLeuArgG1YSerIleLeu 179
 RESULT 2
 AAB87344
 ID AAB87344 standard; protein; 215 AA.

XX AC AAB87344;
 XX AC 22-MAY-2001 (first entry)
 XX DE Human gene 3 encoded secreted protein HMT078, SEQ ID NO:85.
 XX KW Human; secreted protein; proliferative disorder; cancer; tumour;
 XX KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 XX KW immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;
 XX KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 XX KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 XX KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 XX KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 XX KW gastrointestinal disorder; pregnancy-related disorder;
 XX KW endocrine disorder; infection; wound healing; vlnarary; cell culture;
 XX KW chemotaxis; food additive; binding partner identification.
 XX OS Homo sapiens.
 XX PN MO200118022-A1.
 XX PD 15-MAR-2001.
 XX PF 31-AUG-2000; 2000MC-US024008.
 XX PR 03-SEP-1999; 99US-0152315P.
 XX PR 03-SEP-1999; 99US-0152317P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX NI N1 J, Baker KP, Biese CE, Flacella M, Komatsoulis GA, Rosen CA;
 XX PI Soppet DR, Young PE, Edner R, Duan DR, Olsen HS, Lafleur DW;
 XX PI Moore PA, Shi Y, Wei Y, Florence KA;
 XX WPI: 2001-203081/20.
 XX DR N-PSDB; AAF91860.
 XX PT Nucleic acid molecules encoding human secreted proteins, used in
 XX PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 XX PT Parkinson's diseases and cancers.
 XX PS Claim 11, Page 532-533; 607pp; English.
 XX AA AAF91858- AAF91929 represent cDNAs corresponding to 52 human secreted
 XX CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
 XX CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
 XX CC and their corresponding secreted proteins are useful for preventing,
 XX CC treating or ameliorating medical conditions, e.g., by protein or gene
 XX CC therapy. Pathological conditions can be diagnosed by determining the
 XX CC amount of the new protein in a sample or by determining the presence of
 XX CC mutations in the new genes. Specific uses are described for each of the
 XX CC 52 genes, based on the tissues in which they are most highly expressed,
 XX CC and include developing products for the diagnosis or treatment of
 XX CC proliferative disorders, cancer, tumours, foetal and developmental
 XX CC abnormalities, haematopoietic disorders, diseases of the immune system,
 XX CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 XX CC allergies, neurological disorders (e.g., Alzheimer's disease, skin
 XX CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
 XX CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 XX CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 XX CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 XX CC disorders, and infections. The proteins can also be used to aid wound
 XX CC healing and epithelial cell proliferation, to prevent skin aging due to
 XX CC sunburn, to maintain organs before transplantation, for supporting cell
 XX CC culture of primary tissues, to regenerate tissues, to identify their
 XX CC cognate ligands or binding partners, and in chemotaxis, and can be used
 XX CC as a food additive or preservative to modify storage properties.
 XX CC Antibodies specific for a protein of the invention can be used in
 XX CC alleviating symptoms associated with the disorders mentioned above, and
 XX CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 XX CC immunosorbent assay (ELISA). The present sequence represents a human
 XX CC secreted protein of the invention

XX SQ Sequence 215 AA;
Alignment Scores:
Pred. No.: 4,68e-69 Length: 215
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 4 Gaps: 0
US-10-699-035a-1 (1-537) x AAB87344 (1-215)
QY 1 GGAGACCTGATGTTCTGCTGAGACAGCTACAGCCAGCTCTCTACATACAGTTCTCCCG 60
DB 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGlnPheSerArg 52
QY 61 GTTCCGGAGTTGTGGGGCAGCTGTGGCTCCACTGCCCTGGGGCACCGGGGCTTGGCGT 120
DB 53 ValArgGlnPheValGlnGlnLeuValAlaProLeuProLeuGlnThrGlnAlaLeuArg 72
QY 121 GCAGCTGTGACGTGGGGAGTGGGCGCATACAGCGAGTCCCTTGGCGCACGACAG 180
DB 73 AlaSerLeuValHisValGlnSerArgProTyrThrGlnPheProPheGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGAGATGCGGTGCTTCTGCCAGCGCATGGGTGACACCCAC 240
DB 93 SerGlnGlnAlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlnAspThrHis 112
QY 241 ACTGGCTGCGCTGTGCTATGCTCAAGAAACAGCTGTTGTGAAGCATCAGGTGCCCG 300
DB 113 ThrGlnLeuAlaLeuValTyrAlaTyrGlnGlnLeuPheAlaGlnAlaSerGlnAlaArg 132
QY 301 CCAGGGGTGCGCCAAAGTGTGCTGTGGGTGACAGATGGGGCTCCAGCGACCTGAGGGC 360
DB 133 ProGlnValProTyrValLeuValIrrValThrAspGlnGlnSerSerAspProValGln 152
QY 361 CCCCCCATGAGAGCTCAAGACCTGGGCTGACCGCTTCACTTGTCAAGACCGCGCA 420
DB 153 ProPheMetGlnGlnLeuLeuAspLeuGlnValThrValPheLeuValSerThrGlnArg 172
QY 421 GCGCACTTCTGTGAGCTGTGACCGCTGCTCCCTGAGCCCTGCGAGAAACCTGCACTTT 480
DB 173 GlyAspPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 192
QY 481 GTGAGCTGTGAGTGCATCATTCATTCCTCAAGAGCTGAGGGGCTCCATTCCTC 534
DB 193 ValAspValAspAspLeuHisLeuValGlnGlnLeuArgGlnSerLeuLeu 210
RESULT 3
ID ABB65347 standard; protein; 215 AA.
AC ABB65347;
XX 27-AUG-2002 (first entry)
XX Human albumin fusion protein #2022.
DE
XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW human serum albumin; HSA; cancer; reproductive disorder;
KW digestive disorder; immune disorder; endocrine disorder;
KW haematopoietic disorder; neural disorder; connective disorder;
KW cytostatic; antifertility; antiinflammatory; anticancer;
KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW osteopathic; antiarthritic.
OS Homo sapiens.
OS Synthetic.
XX WO200177137-A1.
XX

PD 18-OCT-2001.
XX 12-APR-2001; 2001WO-US011988.
XX 12-APR-2000; 2000US-0229358P.
XX 25-APR-2000; 2000US-0199384P.
XX 21-DEC-2000; 2000US-0256931P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Haseltine WA;
XX WPI; 2002-010886/01.
XX
XX New fusion protein for treating disease e.g. diabetes comprises an
XX albumin fused to a therapeutic protein.
XX
XX Claim 1; Page 1935; 2102pp; English.
XX
XX The present invention relates to albumin fusion proteins comprising a
XX therapeutic protein X and human albumin (HA, also known as human serum
XX albumin, HSA). The proteins are useful for treating a disease or disorder
XX that may be modulated by therapeutic protein X. The albumin extends the
XX shelf-life of protein X, and may increase its biological in vitro/in vivo
XX activity. The protein is useful for treating and diagnosing disorders
XX such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
XX disease, ulcerative colitis), immune disorders (e.g. acquired
XX immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
XX haematopoietic disorders, neural disorders (e.g. Alzheimer's,
XX Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
XX schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).
XX ABB63326-ABB65518 represent albumin fusion proteins of the invention
XX
SQ Sequence 215 AA;
Alignment Scores:
Pred. No.: 4,68e-69 Length: 215
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 5 Gaps: 0
US-10-699-035a-1 (1-537) x ABB65347 (1-215)
QY 1 GGAGACCTGATGTTCTGCTGAGACAGCTACAGCCAGCTCTCTACATACAGTTCTCCCG 60
DB 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGlnPheSerArg 52
QY 61 GTTCCGGAGTTGTGGGGCAGCTGTGGCTCCACTGCCCTGGGGCACCGGGGCTTGGCGT 120
DB 53 ValArgGlnPheValGlnGlnLeuValAlaProLeuProLeuGlnThrGlnAlaLeuArg 72
QY 121 GCAGCTGTGACGTGGGGAGTGGGCGCATACAGCGAGTCCCTTGGCGCACGACAG 180
DB 73 AlaSerLeuValHisValGlnSerArgProTyrThrGlnPheProPheGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGAGATGCGGTGCTTCTGCCAGCGCATGGGTGACACCCAC 240
DB 93 SerGlnGlnAlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlnAspThrHis 112
QY 241 ACTGGCTGCGCTGTGCTATGCTCAAGAAACAGCTGTTGTGAAGCATCAGGTGCCCG 300
DB 113 ThrGlnLeuAlaLeuValTyrAlaTyrGlnGlnLeuPheAlaGlnAlaSerGlnAlaArg 132
QY 301 CCAGGGGTGCGCCAAAGTGTGCTGTGGGTGACAGATGGGGCTCCAGCGACCTGTTGGC 360
DB 133 ProGlnValProTyrValLeuValIrrValThrAspGlnGlnSerSerAspProValGln 152
QY 361 CCCCCCATGAGAGCTCAAGACCTGGGCTGACCGTGTTCATTCATTCCTCAAGAGCTG 420
DB 153 ProPheMetGlnGlnLeuLeuAspLeuGlnValThrValPheLeuValSerThrGlnArg 172

QY 421 GGCACTTCCTGGAGCTGTCAAGCGCTGCTCAGCGCCCTCCGAGAGACACTGCACTTT 480
CC |||||
CC 173 GlyanPheLeuGluLeuSerAlaIalaSerAlaProAlaGluIuYHHisIleuHisPhe 192
CC |||||
QY 481 GTGACGCGATGACCTGCACATCATTTGCCAAGAGCTGAGGGGCTCCATTCTC 534
CC |||||
CC 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210
CC |||||
RESULT 4
ADL78614
ID ADL78614 standard; protein; 215 AA.
AC ADL78614;
XX 20-MAY-2004 (first entry)
DT
XX
XX Albumin fusion protein related therapeutic protein X, SEQ ID NO 2096.
DE
XX albumin fusion protein; cytosstatic; antihaemic; antiarthritic;
KM antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
KM antiapoptotic; antibacterial; osteopathic; dermatological; anti-gout;
KM immunomodulator; antiarrhythmic; cardiac; neurotropic; antidiabetic;
KM nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
KM antidiabetic; anabolic; hypertensive; vulnerary; gene therapy; cancer;
KM reproductive system disorder; therapeutic protein.
XX
XX Unidentified.
OS
XX US2004010134-A1.
PN
XX 15-JAN-2004.
PD
XX 12-APR-2001; 2001US-00833245.
PF
XX 12-APR-2000; 2000US-0229358P.
PR 25-APR-2000; 2000US-0199384P.
PR 21-DEC-2000; 2000US-0256931P.
XX
XX (ROSE/) ROSEN C A.
PA (HASE/) HASELTINE W A.
XX
XX Rosen CA, Haseeltine WA;
PI
XX WPI; 2004-090519/09.
DR
XX
XX New albumin fusion proteins, useful for diagnosing, treating, preventing
PT or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,
PT asthma, inflammatory bowel disease or Alzheimer's disease.
PT
XX
XX Disclosure; SEQ ID NO 2096; 279pp; English.
PS
XX
XX The invention relates to a novel albumin fusion protein. The invention
CC further relates to: a composition comprising the albumin fusion protein
CC and a pharmaceutical carrier; a kit comprising the composition of the
CC albumin fusion protein formula; a method of treating a disease or
CC disorder in a patient comprising the step of administering the albumin
CC fusion protein; a method of treating a patient with a disease or disorder
CC that is modulated by Therapeutic protein: X, or its fragment or variant;
CC a method of extending the shelf life of Therapeutic protein: X, or its
CC fragment or variant; a nucleic acid molecule comprising a polynucleotide
CC sequence encoding the albumin fusion protein; a vector comprising the
CC nucleic acid molecule of the albumin fusion protein; and a host cell
CC comprising the nucleic acid molecule of the albumin fusion protein. The
CC albumin fusion protein and its compositions have the following
CC activities: cytostatic, antinaemic, antiarthritic, antiasthmatic, anti-
CC HIV, immunosuppressive, antiinflammatory, antipsoptic, antibacterial,
CC osteopathic, dermatological, anti-gout, immunomodulator, antiarrhythmic,
CC neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic,
CC hypertensive, and vulnerary. The albumin fusion protein nucleic acid may
CC be used in gene therapy to treat disorders. The albumin fusion protein is
CC useful for diagnosing, treating, preventing or ameliorating diseases or
CC disorders comprising indication: Y. The diseases or disorders include:

CC cancer (e.g. leukemia, colon, bone, breast, liver or lung cancer),
CC immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute
CC lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,
CC autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
CC disease), reproductive system disorders (e.g. prostatitis, inguinal
CC hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
CC Leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,
CC Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
CC or cachexia), cardiovascular disease (e.g. rhabdomyomas, heart disease,
CC arrhythmia, cardiac arrest, heat valve disease, hypernatraemia or
CC hyponatremia), mixed foetal diseases (e.g. foetal alcohol syndrome,
CC Down's syndrome, Patou syndrome, Turner's syndrome, Apert syndrome or Tay-
CC Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
CC tract infections or renal disorders), neural or sensory disease (e.g.
CC Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
CC cerebellar ataxia, attention deficit disorder, autism or obsessive
CC compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
CC occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
CC disease or glomerulonephritis), digestive diseases (e.g. portal
CC hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
CC or connective tissue or epithelial diseases (e.g. Crohn's disease,
CC scleroderma, wound healing or epidermolysis bullosa). This sequence
CC represents a therapeutic protein X relating to the albumin fusion protein
CC of the invention. The sequence listing data for this specification was
CC downloaded from the USPTO website.
XX
XX
SQ Sequence 215 AA;
Alignment Scores:
Pred. No.: 4,68e-69 Length: 215
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
Gaps: 0
US-10-699-035A-1 (1-537) x ADL78614 (1-215)
QY 1 GGAGCACTGATGTTCTCTGCTGAGACGCTCAGCCAGCGTCTCTACTAGCACTTCCCGG 60
CC |||||
DB 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisIleYrGluPheSerArg 52
CC |||||
QY 61 GTTCGGAGATTGTGGGGCAGCTGGTGCCTCACTGCCCTGGGCAGCCGGGCTTCGCT 120
CC |||||
DB 53 ValArgGluPheValIleGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
CC |||||
QY 121 GGCACTGCTGTCAGCTGGGAGATCGGCATACACCCAGTCCCTTGGGCAGACAGC 180
CC |||||
DB 73 AlaSerLeuValHisValGlySerArgProIyrThrGluPheProPheGlyGlnHisSer 92
CC |||||
QY 181 TCGGGTGAAGCTGCGCCAGAGATCGGAGTGGTCTTCTGCCAGCCATGGGTGACCCGAC 240
CC |||||
DB 93 SerGlyGluAlaIalaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlyAspThrHis 112
CC |||||
QY 241 ACTGGCTGGCGCTGTATGCCAAGAGACAGCTGTTTGTGTGAAGCATCAGGTGCCGG 300
CC |||||
DB 113 ThrGlyLeuAlaLeuValIyrAlaIleGlnGlnLeuPheAlaGlnAlaSerGlyAlaArg 132
CC |||||
QY 301 CCAGGGGTGCCCAAGTCTGTGTGGTGAACAGATGGCGGCTCCAGGACCTGTGGGC 360
CC |||||
DB 133 ProGlyValProIyrValLeuValIyrAlaThrAspGlyGlySerSerAspProValGly 152
CC |||||
QY 361 CCCCCATGAGAGAGCTCAAGACCTGGGCGCTCAACCGGTTCATTGTGAGACCGGCGGA 420
CC |||||
DB 153 ProPheMetGlnGluLeuLeuAspLeuGlyAlaIleValIleValSerThrGlyArg 172
CC |||||
QY 421 GGCACTTCCTGGAAGCTGTCAAGCGCTGCTCAGCGCCCTCCGAGAGACACTGCACTTT 480
CC |||||
DB 173 GlyanPheLeuGluLeuSerAlaIalaSerAlaProAlaGluIuYHHisIleuHisPhe 192
CC |||||
QY 481 GTGACGCGATGACCTGCACATCATTTGCCAAGAGCTGAGGGGCTCCATTCTC 534
CC |||||
DB 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210
CC |||||

RESULT 5
AAB87424
ID AAB87424 standard; protein; 226 AA.
XX
AC AAB87424;
XX
DT 22-MAY-2001 (first entry)
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
OS Homo sapiens.
XX
PN MO200118022-A1.
XX
PD 15-MAR-2001.
XX
PF 31-AUG-2000; 2000MO-US024008.
XX
PR 03-SEP-1999; 99US-0152315P.
PR 03-SEP-1999; 99US-0152317P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
PI Soppet DR, Young PE, Edner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
DR WPI; 2001-203081/20.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
PT
XX
PS Disclosure; Page 18; 607p; English.
XX
CC AAF1856-AAF1929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB8742-AA87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiotensin disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and

CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
SQ Sequence 226 AA;
XX
Alignment Scores:
Pred. No.: 4,73e-69 Length: 226
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 4 Gaps: 0
US-10-699-035A-1 (1-537) x AAB87424 (1-226)
QY 1 GGGAGCCGATGTTCTCTGAGAGCTCAGCCAGCCGCTCTCAATGAGTTCCCGG 60
DB 46 GYAAPLeuMetPheLeuLeuAspSerSerAlaSerValSerHisrGlupheserArg 65
QY 61 GTTCGGAGTTTGTGGGACAGCTGGGCTCCAGTCCCTGGGACCGGGGCTGGCGT 120
DB 66 VALArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 85
QY 121 GCCAGCTGTGTGACAGTGGGAGTGGCCATACACCGAGTTCCCTGGCCAGACAGC 180
DB 86 AlaserLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 105
QY 181 TCGGTGAGGCTGCCAGAGATGCGGTGCTTGTGCCAGCGATGGGTGACACCCAG 240
DB 106 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 125
QY 241 ACTGCGCTGGCGCTGTATGTCAGAGAACAGCTGTTGTGAAGCATCAGAGTCCCGG 300
DB 126 ThrGlyLeuAlaLeuValTyrAlaGlySerGlnLeuPheAlaGlnHisSerGlyAlaArg 145
QY 301 CCAGGCTGGCCCAAGAGTCTGTGTGGTGTGACAGATGGCGGCTCCAGGACCTGTGGGC 360
DB 146 ProGlyValProGlyValLeuValTyrValThrAspGlyGlySerSerAspProValGly 165
QY 361 CCCCCAGTGGAGGCTCAAGGACCTGGGCGCTCACCGTTCATGTGTGACAGCCGCGGA 420
DB 166 ProPheMetGlnGluLeuLeuAspLeuGlyValThrValPheIleValSerThrGlyArg 185
QY 421 GCGAATCTCTGGAGCTGTACAGCCGCTCTCAGCGCCCTGGCGAGAACACCTGCATT 480
DB 186 GlyAsnPheLeuGluLeuSerHisAlaAlaSerHisAlaProAlaGlnHisSerHisPhe 205
QY 481 GTGACGTGATGATGATCAGATCATCTTGTCCAGAGGAGGCTCCATTCTC 534
DB 206 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 223
RESULT 6
AAB87418
ID AAB87418 standard; protein; 242 AA.
XX
AC AAB87418;
XX
DT 22-MAY-2001 (first entry)
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX

OS Homo sapiens.
 XX WO200118022-A1.
 XX 15-MAR-2001.
 XX 31-AUG-2000; 2000WO-US024008.
 XX 03-SEP-1999; 99US-0152315P.
 PR 03-SEP-1999; 99US-0152317P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
 PI Sopet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
 PI Moore PA, Shi Y, Wei Y, Florence KA;
 XX WPI; 2001-203081/20.
 DR Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers.
 PS Disclosure: Page 18; 607bp; English.
 XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
 CC protein genes, and AAB87342-AA87413 represent the proteins they encode.
 CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 52 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
 CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunoassorbent assay (ELISA). The present sequence represents a human
 CC secreted protein fragment referred to in the disclosure of the invention
 XX
 XX SQ Sequence 242 AA;
 Alignment Scores:
 Pred. No.: 4 79e-69 Length: 242
 Score: 902.00 Matches: 178
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 4 Gaps: 0
 US-10-699-035A-1 (1-537) x AAB87418 (1-242)
 QY 1 GGGGACCTGATGTTCTGCTGGACAGCTGACGAGCTCTCTACATACAGTTCTCCCG 60
 DB 60 GtlyAspIeueMetPheIeueIueNspSersSerIaSerValSerHisrlyIuIuHeserArg 79
 QY 61 GTTCGGAGATTGTGTGGGGCAAGCTGTGCTCCACTGCCCCCTGGGCAACCGGGGCTTGCGT 120

DB 80 ValArgIuIuPheValGIgYIuIeueValAlaProIeueProIeueGIYThrGIYAlaIeueArg 99
 QY 121 GCCAGCTGTGTGCAAGTGGGCAAGTCGGCCATACACCGAGTCCCTTGGGCAACAGC 180
 DB 100 AlaSerIeueValHisValGIgYSerArgProIYThrGIuPheProPheGIYIuHisSer 119
 QY 181 TCGGGGTAGGCTGCCCAAGATGCGGCTGTCTTCTGCGCCAGCGCATGGGTGACACCCAC 240
 DB 120 SerGIYIuAlaIaGIuIaSerPalValArgAlaSerIaGIuIaArgMetGIYAspThrHis 139
 QY 241 ACTGGCCCTGGCGCTGCTAGTATAGCCAGGAACAGCGTGTTCGTAAGCATCAGTGCCCGG 300
 DB 140 ThrGIYIeueAlaIeueValTYrAlaIuSerGIuIeuePheAlaGIuAlaSerGIYAlaArg 159
 QY 301 CCAGGCGTGGCCAAAGT 360
 DB 160 ProGIYValProIYsValIeueValTYrValThrAspGIYIuSerSerAspProValGIY 179
 QY 361 CCCCCATGCAAGAGCTCAAGACCTGGGCGTCAACCGTTCATTGTTCAGCACCGGCCGA 420
 DB 180 ProPheMetGIuGIuIeueIuYAspIeueGIYValThrValPheIleValSerThrGIYArg 199
 QY 421 GGGCACTTCCTGGAGCTGTGACCGCGCTGCTGACCGCCCTGCGGAGAGCACCTGCACTT 480
 DB 200 GtlyAspPheIeueGIuIeueSerAlaIaIaIaSerAlaProIaGIuIuYHisIeueHisPhe 219
 QY 481 GTGCACTGTGATGACCTGTGACATGATTTGTCCAGAGCTGAGGAGCTCCATTCTC 534
 DB 220 ValAspValaAspPheIuHisIleIleValGIuGIuIeueArgGIYSerIleIeue 237
 RESULT 7
 AAB88340
 ID AAB88340 standard; protein, 285 AA.
 AC AAB88340;
 DT 23-MAY-2001 (first entry)
 XX Human membrane or secretory protein clone PSEC0053.
 DE Human membrane or secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes.
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 OS Homo sapiens.
 XX EP1067182-A2.
 PN 10-JAN-2001.
 PD 07-JUL-2000; 2000EP-00114090.
 PF 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 PI WPI; 2001-093889/11.
 DR N-PSDB; AAF93767.
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development.
 XX Claim 1; SEQ ID NO 48; 609BP + Sequence Listing; English.
 PS This invention relates to nucleic acid sequences AAF93744 - AAF93916
 XX which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF92232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can

be used in vaccines. The polynucleotide sequences can be used in gene therapy. The polynucleotide sequences and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptide in samples (e.g. by enzyme linked immunosorbent assay (ELISA). Examples of diseases which may be treated include rheumatoid arthritis and diabetes

XX Sequence 285 AA;

Alignment Scores:

Pred. No.:	4,95e-69	Length:	285
Score:	902.00	Matches:	178
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	86.0%	Indels:	0
DB:	4	Gaps:	0

US-10-699-035a-1 (1-537) x AAB88340 (1-285)

```

QY 1 GGGAGCTGATGTTCTGCTGAGACGCTCAGCCAGCGTCTTCTACAGATTCTCCCG 60
DB 33 GYAAPLeuMePheLeuLeuSPSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCCGAGATTGTTGGGGAGCGTGGGCTCCACCTGCCCTGGGACCGGGCCCTGGCT 120
DB 53 VALARGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY 121 GCCAGTCTGTCACGTCGGGAGTCGGGCATACACCGATTCCCTCGGACGACACAGC 180
DB 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGAGATGCGGTGCGTCTTCTGCCAGCGCATGGGTGAACCCAC 240
DB 93 SerGlyGluAlaAlaGlnPheAlaValArgAlaSerAlaGlnArgMetGlyPheThrHis 112
QY 241 ACTGGCCTGCGCGCTGCTCTATGCAAGAAACAGCTGTTGTCGAGCATCAGTCCCGG 300
DB 113 ThrGlyLeuAlaLeuValTyrAlaTyrGlnGlnLeuPheAlaGlnAlaSerGlyAlaArg 132
QY 301 CCAGGGGATGCCCAAGATGCTGGGTGTCAGATGGGCGCTCCAGCGACCTGTGGAGC 360
DB 133 ProGlyValProLysValLeuValTyrValThrValPheGlyGlySerSerSerProValGly 152
QY 361 CCCCCCATGACAGAGCTCAAGACCTGGGCGTCAACCGTTCATTGTCAAGACCGGCGCA 420
DB 153 ProPheMetGlnGluLeuLysPheGlyValThrValPheIleValSerThrGlyArg 172
QY 421 GGCACCTTCTGAGAGTGTCAAGCGGCTGCTCAGCCCTCGGACGACGACCTGCACTT 480
DB 173 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 192
QY 481 GTGAGAGTGAATGACCTGACATCATTTGCAAGAGCTGAGGGGCTCCATTCTC 534
DB 193 ValAspValAspPheLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210

```

RESULT 8

ADY63045

ID ADY63045 standard; protein; 285 AA.

XX ADY63045;

DT 02-JUN-2005 (first entry)

```

XX XX Human clone PSEC0053 protein, SEQ ID 48.
DE DE
XX XX Gene therapy.
KW KW
XX XX Homo sapiens.
OS OS
XX XX EP1514933-A1.
PN PN
XX XX 16-MAR-2005.
PD PD
XX XX 07-JUL-2000; 2004EP-00027228.
PF PF
XX XX 08-JUL-1999; 99JP-00194179.
PR PR
XX XX 11-JAN-2000; 2000JP-00118775.
PR PR
XX XX 02-MAY-2000; 2000JP-00183766.
PR PR
XX XX 07-JUL-2000; 2000EP-00114090.
XX XX
XX XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX XX
XX XX Ota T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI PI
XX XX MPI: 2005-203865/22.
DR DR
XX XX N-PSDB; ADY63044.
DR DR
XX XX Novel isolated polynucleotide encoding human secretory proteins or
PT PT membrane proteins, useful for examination and diagnosis of abnormality of
PT PT human secretory proteins.
XX XX
XX XX Disclosure; SEQ ID NO 48; 1240pp; English.
PS PS
XX XX The present invention relates to novel human secretory proteins or
CC CC membrane proteins, and their coding sequences. The present sequence is
CC CC one such protein sequence. The coding sequences of the invention are
CC CC useful for examination and diagnosis of abnormality of the human
CC CC secretory proteins and in gene therapy methods. The coding sequences and
CC CC proteins are useful as candidates for medicines or as target molecules
CC CC for developing medicines. Antibodies against the proteins of the
CC CC invention are useful for treating diseases that are associated with the
CC CC proteins. Note: The sequence data for this patent did not form part of
CC CC the printed specification, but was obtained from sequence information
XX XX supplied by the European Patent Office.
XX XX
XX XX Sequence 285 AA;
SQ SQ

```

Alignment Scores:

Pred. No.:	4,95e-69	Length:	285
Score:	902.00 <td>Matches:</td> <td>178</td>	Matches:	178
Percent Similarity:	100.0% <td>Conservative:</td> <td>0</td>	Conservative:	0
Best Local Similarity:	100.0% <td>Mismatches:</td> <td>0</td>	Mismatches:	0
Query Match:	86.0% <td>Indels:</td> <td>0</td>	Indels:	0
DB:	9	Gaps:	0

US-10-699-035a-1 (1-537) x ADY63045 (1-285)

```

QY 1 GGGAGCTGATGTTCTGCTGAGACGCTCAGCCAGCGTCTTCTACAGATTCTCCCG 60
DB 33 GYAAPLeuMePheLeuLeuSPSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCCGAGATTGTTGGGGAGCGTGGGCTCCACCTGCCCTGGGACCGGGCCCTGGCT 120
DB 53 VALARGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY 121 GCCAGTCTGTCACGTCGGGAGTCGGGCATACACCGATTCCCTCGGACGACGACCTGCACTT 180
DB 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGAGATGCGGTGCGTCTTCTGCCAGCGCATGGGTGAACCCAC 240
DB 93 SerGlyGluAlaAlaGlnPheAlaValArgAlaSerAlaGlnArgMetGlyPheThrHis 112
QY 241 ACTGGCCTGCGCGCTGCTCTATGCAAGAAACAGCTGTTGTCGAGCATCAGTCCCGG 300

```

Db 113 ThrGlyLeuAlaLeuValTyrAlaLysGluGlnLeuPheAlaGluAlaSerGlyAlaArg 132
Qy 301 CCAAGGGGTGCCCAAGTCTGGTGTGGGTGAACAATGGCGGCTCCAGGACCTGTGGGC 360
Db 133 ProGlyValProlYsValLeuValTrrValThraSpGlyGlySerSerAspProValGly 152
Qy 361 CCCCCATGACAGAGACTCAAGACCTGGCGCTGACCCGTTCATTGTACGACCCGGCGGA 420
Db 153 ProPmetGInGluLeuLysAspLeuGlyValThrrValPheIleValSerThrGlyArg 172
Qy 421 GGCAACTTCCTGGAGCTGTGACGCGCTGCGCCCGCCGAGAGACCTGCACCTT 480
Db 173 GlyAsnPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGluLysIleuHisPhe 192
Qy 481 GTGACGTGATGACCTGCACATCATTTGCAAGAGCTGAGGGGCTCCATTCTC 534
Db 193 ValAspValAspAspLeuHisIleIleValGlnGlnLeuArgGlySerIleLeu 210
RESULT 9
AAE32502
ID AAE32502 standard; protein; 418 AA.
AC AAE32502;
XX 24-MAR-2003 (first entry)
XX Human Willebrand Factor A domain related-protein (WARP).
XX Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
KW extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
KW gene therapy; human.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT Protein /label= Signal_peptide
FT 19..418
FT /note= "Human mature WARP protein"
FT Modified-site 148
FT /note= "O-glycosylation site"
FT Misc-difference 210..211
FT /note= "Encoded by CTCGGC"
FT Modified-site 264
FT /note= "N-glycosylation site"
FT Modified-site 359
FT /note= "N-glycosylation site"
FT Modified-site 361
FT /note= "O-glycosylation site"
FT Disulfide-bond 369..393
XX
XX MO200288184-A1.
XX
XX 07-NOV-2002.
XX
XX 02-MAY-2002; 2002WO-AU000542.
XX
XX 02-MAY-2001; 2001AU-00004701.
XX
XX (MURD-) MURDOCH CHILDRENS RES INST.
XX
XX Bateman JF, Fitzgerald DJ;
XX
XX MPI; 2003-111873/10.
XX
XX N-PSDB; AAD50399.
XX
XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
XX the manufacture of a medicament in the treatment of a disease condition
XX of the extracellular matrix, in particular arthritis.
XX
XX Claim 11; Page 76-78; 103pp; English.
XX
XX The invention relates to Willebrand Factor A domain related-protein

CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a
CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medicament in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
XX present sequence is human WARP protein
XX
SQ Sequence 418 AA:
XX
XX
XX Alignment Scores:
XX Pred. No.: 5,35e-69 Length: 418
XX Score: 902.00 Matches: 178
XX Percent Similarity: 100.0% Conservative: 0
XX Best Local Similarity: 100.0% Mismatches: 0
XX Query Match: 86.0% Indels: 0
XX DB: 6 Gaps: 0
XX
XX US-10-699-035A-1 (1-537) x AAE32502 (1-418)
Qy 1 GGGGACCTGATGCTCTGCTGGAAGCTGACGCGCTCTCACTACGAGTTCTCCCGG 60
Db 33 GlyAspLeuValPheLeuLysAspSerSerAlaSerValSerHisTyrGlnPheSerArg 52
Qy 61 GTTCGGAGATTGTGGGACAGCTGTGGCTTCACCTGCGCCCTGGACCGGCGCTGGCT 120
Db 53 ValArgGlnPheValGlnGlnLeuValAlaProLeuProLeuGlyThrGlnAlaLeuArg 72
Qy 121 GCCAGTGTGTGACAGTGGGAGTGGGCGCATACACCGAGTTCCCTTGGGCGACACAGC 180
Db 73 AlaSerLeuValHisValGlySerArgProTyrThrGlnPheProPheGlyGlnHisSer 92
Qy 181 TCGGTGAGGCTGCCGAGATGGCGGTGTCTTCTGCCGCGGACGAGGAGACACCCAC 240
Db 93 SerGlyGlnAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
Qy 241 ACTGGCTGCGCGCTGTCTATGCCAAGAACAGCTGTTTGCTGAAGCATCAGTGGCCGG 300
Db 113 ThrGlyLeuAlaLeuValTyrAlaLysGlnGlnLeuPheAlaGlnAlaSerGlyAlaArg 132
Qy 301 CCAAGGGGTGCCCAAGTCTGGTGTGGGTGAACAATGGCGGCTCCAGGACCTGTGGGC 360
Db 133 ProGlyValProlYsValLeuValTrrValThraSpGlyGlySerSerAspProValGly 152
Qy 361 CCCCCATGACAGAGACTCAAGACCTGGCGCTGACCCGTTCATTGTACGACCCGGCGGA 420
Db 153 ProPmetGInGluLeuLysAspLeuGlyValThrrValPheIleValSerThrGlyArg 172
Qy 421 GGCAACTTCCTGGAGCTGTGACGCGCTGCGCCCGCCGAGAGACCTGCACCTT 480
Db 173 GlyAsnPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGluLysIleuHisPhe 192
Qy 481 GTGACGTGATGACCTGCACATCATTTGCAAGAGCTGAGGGGCTCCATTCTC 534
Db 193 ValAspValAspAspLeuHisIleIleValGlnGlnLeuArgGlySerIleLeu 210
RESULT 10
ABP69674
ID ABP69674 standard; protein; 445 AA.
XX
XX ABP69674;
XX
XX 20-JAN-2003 (first entry)
XX
XX Human polypeptide SEQ ID NO 1721.
XX
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
XX cell-proliferative disorder; neurodegenerative disease; bacterial;
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW

KM arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
 KM antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
 KM haemostatic; vulnerary; fungicide; antibacterial; vincible; protozoacide
 antiarthritic.

OS Homo sapiens.

PN WO200270539-A2.

PD 12-SEP-2002.

PF 05-MAR-2002; 2002WO-US005095

PR 05-MAR-2001; 2001US-00799451.

PA (HYSE-) HYSEQ INC.

P1 Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F
P1 Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
P1 Wehrman T, Wang J, Wang D, Drmanac RT;

DR WPI; 2002-759812/82.
DR N-PSDB; ABZ11891.

DR N-PSDB; ABZ11891.

PT New polynucleotides comprising sequences assembled from expressed PT sequence tags (ESTs), useful for treating cell-proliferative, neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet or coagulation disorders.

PS Claim 9; SEQ ID NO 1721; 1012pp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide (1) comprising a
CC nucleotide sequence selected from any of 948 sequences (ABZ11119-
CC ABZ12066) or their mature protein coding portion, active domain coding
CC protein or complementary sequences. The polynucleotides are useful for
CC identifying expressed genes or for physical mapping of human genome. The
CC encoded polypeptides (ABP689802-ABP69849) are useful as molecular weight
CC markers, as a food supplement, for generating antibodies, in medical
CC imaging, screening and diagnostic assays and for treating cell-
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

SQ Sequence 445 AA;

Alignment Scores:

Pred. No.:	445	length:	1788
Score:	9.01e-9	Matches:	1788
Percent Similarity:	100.0%	Conservative:	0
Best local Similarity:	100.0%	Mismatches:	0
Query Match:	86.0%	Indels:	0
DB:	5	Gaps:	0

US-10-699-035A-1 (1-537) X ABP69674 (1-445)

Qy	1	GGGAGCTATAGTTCCTGCTGGACAGCTCAGCAGGCTCTCACTGAGATTTCTCCGG	60
Db	33	GLIAspneuetherheuleuulaspserSerIalaserValserNiaTyrgLuphSerArg	52
Qy	61	GTTTCGGAGATTGTGGGAGCAGTGGAGGCTCCACTGAGCCCTGGGACAGCGGGGCTTCGT	120
Db	53	ValArgLupheValGlyGluLeuValAlaProLeuProLeuGlyTyrGlyAlaLeuArg	72
Qy	121	GCCAGTCTGGTGCACGTCGGAGTCGGCCATACACGAGTTCCCTTCGGCCAGCAGCAGC	180
Db	73	AlaserLeuValHisValGlySerArgProTyrrTyrGluPheProPheGlyGluHisSer	92
Qy	181	TCGGGTGAGGCTGCCAGAGATGCGTGCCTCTTTCGCCACGCGAGCTGACACCCAC	240

Db	93	SeGcIyGlnIaIaIaGlnhspAlaValaIrgAlaSerAlaGlnaGmGcIyAaerThhIe	112
Qy	241	ACTGAGCTGGAGCGTGTCTATGSCCAAGGAACAGCTGTTGTGTGAAGCATCAGGTGCCGG	300
Db	113	ThrGIyLeuAlaIeuValaIyrgAlaIyGcIuGlnIeuPheAlaGlnaIaSerGIyAlaIaArg	132
Qy	301	CCAGGGGGTCCCAAAAGTGTGTGTGGGTGACAAATGGCGGCTCCACGAACTCGTGGGC	360
Db	133	ProGIyValaProIyValaIeuValaITrpaIAlhThraSpGIyGlySerSetaAepProValaGly	152
Qy	361	CCCCCATCAGAGAGCTCAAGAGACTGGGCGCTCAACCGTTCATTTGTACAGACCGAGCGCA	420
Db	153	ProPheMeGcIuGlnIeuIyehsPheuGlyAlaIThrValaPheIleValaSerThhGIyArg	172
Qy	421	GGCAACTTCCTGAGAGCTGCAGCCGCGTGCCTTACGCCCTGCGAAGAAACATGCACATTT	480
Db	173	GIyAsnPheIuGlnIeuSerAlaIaIaIaSerAlaProAlaGlnIyVhIaIeuIhPhe	192
Qy	481	GTCGAGTGGAGACCTGCACATCTTTCACAAAGTCGAGAGGGCTCATTTCTC	534
Db	193	ValaSpValaAaPspPheuIhIleIleValaGlnGlnIeuAaGgGIySetIleIeu	210

RESULT 11

ID ADH71106 standard; protein; 445 AA

AC ADH71106,

DT 25-MAR-2004 (first entry)

DE Human protein of the invention NOVla SEQ ID NO:2

KM human; cytotactic; immunomodulator; neuroprotective; nootropic;
KM anorectic; antidepressant; antimicrobial; antipneumia; gene therapy;
KM vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease
KM obesity; diabetes; infectious disease; metabolic syndrome X;
KM dyslipidaemia.

OS Homo sapiens.

PN WO2003102155-A2.

PD 11-DEC-2003

PF 03-JUN-2003; 2003WO-US017430

PR	03-JUN-2002	2002US-0385120P.
PR	04-JUN-2002	2002US-0385784P.
PR	05-JUN-2002	2002US-0386041P.
PR	05-JUN-2002	2002US-0386047P.
PR	06-JUN-2002	2002US-0386376P.
PR	06-JUN-2002	2002US-0386453P.
PR	06-JUN-2002	2002US-0386864P.
PR	06-JUN-2002	2002US-0387016P.
PR	07-JUN-2002	2002US-0386796P.
PR	07-JUN-2002	2002US-0386816P.
PR	07-JUN-2002	2002US-0386931P.
PR	07-JUN-2002	2002US-0386942P.
PR	07-JUN-2002	2002US-0386971P.
PR	07-JUN-2002	2002US-0387262P.
PR	08-JUN-2002	2002US-0296960P.
PR	10-JUN-2002	2002US-0387400P.
PR	10-JUN-2002	2002US-0387535P.
PR	11-JUN-2002	2002US-0387610P.
PR	11-JUN-2002	2002US-0387625P.
PR	11-JUN-2002	2002US-0387634P.
PR	11-JUN-2002	2002US-0387688P.
PR	11-JUN-2002	2002US-0387696P.
PR	11-JUN-2002	2002US-0387702P.
PR	11-JUN-2002	2002US-0387836P.
PR	11-JUN-2002	2002US-0387859P.
PR	12-JUN-2002	2002US-0387933P.

PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388012P.
PR 12-JUN-2002; 2002US-0388096P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389118P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389144P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389729P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390068P.
PR 19-JUN-2002; 2002US-0390209P.
PR 21-JUN-2002; 2002US-0390763P.
PR 17-JUL-2002; 2002US-0396706P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402156P.
PR 09-AUG-2002; 2002US-0402256P.
PR 09-AUG-2002; 2002US-0402389P.
PR 12-AUG-2002; 2002US-0402786P.
PR 12-AUG-2002; 2002US-0402816P.
PR 12-AUG-2002; 2002US-0402832P.
PR 13-AUG-2002; 2002US-0403448P.
PR 13-AUG-2002; 2002US-0403459P.
PR 13-AUG-2002; 2002US-0403531P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0403563P.
PR 15-AUG-2002; 2002US-0406317P.
PR 15-AUG-2002; 2002US-0406317P.
PR 26-AUG-2002; 2002US-0406182P.
PR 26-AUG-2002; 2002US-0406355P.
PR 27-AUG-2002; 2002US-0406240P.
PR 12-SEP-2002; 2002US-0410084P.
PR 20-SEP-2002; 2002US-0412528P.
PR 23-SEP-2002; 2002US-0412731P.
PR 30-SEP-2002; 2002US-0414801P.
PR 30-SEP-2002; 2002US-0414839P.
PR 30-SEP-2002; 2002US-0414840P.
PR 30-SEP-2002; 2002US-0414954P.
PR 09-OCT-2002; 2002US-0417186P.
PR 09-OCT-2002; 2002US-0417406P.
PR 23-OCT-2002; 2002US-0420639P.
PR 28-OCT-2002; 2002US-0421156P.
PR 31-OCT-2002; 2002US-0422690P.
PR 01-NOV-2002; 2002US-0423130P.
PR 05-NOV-2002; 2002US-0423798P.
PR 05-NOV-2002; 2002US-0423798P.
PR 12-NOV-2002; 2002US-0425453P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Alsebrook JP, Alvarez E, Anderson DM, Boidog FL, Casman SJ;
PI Caterberg E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
PI Etenberg S, Gargolli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
PI Gusev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, Macdougall JR;
PI MacLachlan T, Malynkar UM, Mezick AJ, Mallet I, Mishra VS;
PI Padigan M, Patirajan M, Pena CE, Peyman JA, Raha D, Raselli L;
PI Rieger DK, Rutherford ME, Sciore P, Shenoy SG, Shimkets RA;
PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;
PI Zhong H;
XX
XX WPI; 2004-081935/08.
DR N-PSDB; ADH71105.
XX
XX New NOVA polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVA-associated disorders, e.g. cancer, diabetes, infection or
PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
XX
XX Example 1; SEQ ID NO 2; 1880pp; English.
PS
XX
XX The invention relates to a novel isolated polypeptide (NOVA). A
CC

CC polypeptide of the invention has cyrostatic, immunomodulator,
CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
CC antihypertensive activity, and may have a use in gene therapy, and as a
CC vaccine. The polypeptides are encoded by NOVA polynucleotides comprising
CC any of the 303 fully defined nucleotide sequences given in the
CC specification. The polypeptide is useful in the manufacture of a
CC medicament for treating a syndrome associated with a human disease. The
CC polypeptide, polynucleotide and antibody are useful in diagnosing,
CC treating or preventing NOVA-associated disorders, e.g. cancer, cachexia,
CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
CC diseases, metabolic syndrome X or dyslipidemias. The nucleic acids are
CC further used as hybridisation probes, in chromosome mapping, tissue
CC typing, preventive medicine, and pharmacogenomics. The present sequence
CC represents a NOVA polypeptide of the invention.
XX
XX SQ Sequence 445 AA;
Alignment Scores:
Pred. No.: 5,41e-69 Length: 445
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 8 Gaps: 0
US-10-699-035A-1 (1-537) x ADH71106 (1-445)
QY 1 GGGAGCTGATGTTCTCTGAGAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCGG 60
DB 33 GlyaspueMetPheIeuLeuAspSerSerIaSerIaSerIaSerIaSerIaSerIa 52
QY 61 GTTGGGAGTTTGGGGGAGCTGGTGGCTCACTCCCTCCGGGACCGGGGCTGGCT 120
DB 53 ValArgIuPheValIglIuIeuValAlaProIeuProIeuIglIuThiGlyAlaIeuArg 72
QY 121 GCGAGTGTGTGACGTGGGAGTGGGAGTGGGACATACCGAGTCCCTTCGGCAGACAGC 180
DB 73 AlaSerIeuValIhIa 92
QY 181 TCGGGTGAAGCTGCGCCAGATGCGGTGCTCTGCGCCAGCCATGGGTGACACCCAC 240
DB 93 SerIgluIa 112
QY 241 ACTGAGCTGGGCTGTGTATGCAAGAGACAGCTGTTGCTGAAGCATCAGTGGCCGG 300
DB 113 ThrIgluIa 132
QY 301 CCGAGGCTGCCAAGCTGT 360
DB 133 ProIgluIa 152
QY 361 CCCCCATGAGAGCTCAAGACCTGGGCTGACCCGTTCATTTGTCAGGACCGCCGA 420
DB 153 ProIrometIuIuIeuIeuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 172
QY 421 GCGAACTCTCTGAGCTGTACGCGCTGCTCAGCCCTGCCGAGAGACCTGACATT 480
DB 173 GlyaspPheIeuIuIeuSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 192
QY 481 GTGAGCGTGAAGTGCCTGACATCATTTGCAAGAGCTGAGGGGCTCATTTCTC 534
DB 193 ValAspValaIa 210
RESULT 12
AAE03654
ID AAE03654 standard; protein, 185 AA.
XX
XX AAE03654;
AC
XX
XX 06-AUG-2001 (first entry)
DT
XX
XX Human extracellular matrix and cell adhesion molecule-18 (XMA-18).
DE
XX

Human; extracellular matrix and cell adhesion molecule; XMD; gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia; Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy; sickle cell anaemia; thalasassaemia; autoimmune disorder; adenocarcinoma; inflammatory disorder; acquired immune deficiency syndrome; AIDS; Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma; Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer; glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis; osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis; infection; cell proliferative disorder; actinic keratosis; myeloma; arteriosclerosis; neotropic; anticonvulsant; antithyroid; nephrotoxic; neuroprotective; dermatological.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FT Region 1..171

XX FT /note= "Collagen glycoprotein precursor"

XX FT 1..170

XX FT Region /note= "Von Willebrand factor domain score"

XX FT 2..15

XX FT Region /note= "Collagen glycoprotein precursor"

XX FT 37..51

XX FT /note= "Collagen glycoprotein precursor"

XX FT 103..111

XX FT /note= "Collagen glycoprotein precursor"

XX PN MO200142285-A2.

XX PD 14-JUN-2001.

XX PF 05-DEC-2000; 2000WO-US032990.

XX PR 10-DEC-1999; 99US-0172852P.

XX PR 16-DEC-1999; 99US-0172354P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C,

XX PI Baughn MR, Lu DM, Shah P, Au-Young J;

XX DR WPI, 2001-381632/40.

XX DR N-PSDB; AAD08062.

XX PT New human extracellular matrix and cell adhesion molecules and

XX PT polynucleotide sequences encoding them, useful for diagnosis, prevention,

XX PT treatment of genetic, autoimmune and cell proliferative disorders.

XX PS Claim 1; Page 111, 135pp; English.

XX XX

CC The present sequence is a human extracellular matrix and cell adhesion

CC molecule (XMD). The XMD is used for screening a compound for

CC effectiveness as an agonist or antagonist of XMD. The identified agonist

CC or antagonist are used for treating a disease or condition associated

CC with decreased or increased expression of functional XMD. The

CC polynucleotides encoding XMD are useful in somatic or germline gene

CC therapy to correct a genetic deficiency, to express a conditional gene

CC lethal gene product and to express a protein which affords protection

CC against intracellular parasites and also for diagnosis of disorders

CC associated with expression of XMD. They are also used for generating

CC hybridisation probes useful in mapping the naturally occurring genomic

CC sequences and to create knock in humanised animals (pigs) or transgenic

CC animals (mice or rats) to model human diseases. Oligonucleotide or longer

CC fragments derived from the polynucleotide sequences may be used as

CC elements on a microarray. Antibodies which specifically bind XMD may be

CC used for the diagnosis of disorders associated with the expression of

CC XMD, or in assays to monitor patients being treated with XMD. Diseases

CC diagnosed, prevented or treated include genetic disorders such as

CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's

CC disease, myotonic dystrophy, sickle cell anaemia, thalasassaemia,

CC autoimmune/inflammatory disorders such as acquired immune deficiency

CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,

CC atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,

CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,

CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,

CC bacterial, fungal, parasitic, protozoal and helminthic infections and

CC cell proliferative disorders such as actinic keratosis, arteriosclerosis

CC and cancer including breast, bladder, bone marrow, brain and uterus

CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma

XX SQ Sequence 185 AA;

XX

XX Alignment Scores:

XX Pred. No.: 1,096-67 Length: 185

XX Score: 886.00 Matches: 175

XX Percent Similarity: 100.0% Conservative: 0

XX Best Local Similarity: 100.0% Mismatches: 0

XX Query Match: 84.5% Indels: 0

XX DB: 4 Gaps: 0

XX

XX US-10-699-035a-1 (1-537) x AAE03654 (1-185)

XX

QY 10 ATGTTCTGTGAGACGCTCAGCCAGCGCTCTCTCACTACGAGTTCTCCGGGTTGGAG 69

DB 1 MetPheLeuLeuAspSerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlu 20

QY 70 TTTGTGGGGCAGCTGTGTGCTTCACTGCCCCCTGGGACCCGGGCTGTGCTGCTG 129

DB 21 PheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuValAlaSerLeu 40

QY 130 GTGCGCGTGGGCGAGCGGCGGCATACACGAGTTCCCTTGGCGAGCAGCTCGGGGTG 189

DB 41 ValHisValGlySerArgProTyrThrGlnPheProPheGlyGlnHisSerSerArgGlu 60

QY 190 GCTGCCAGATGCGCGTGGCTTCTGCGCAGCGCATGGGTGACCCACACTGGGCTG 249

DB 61 AlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeu 80

QY 250 GCGCTGTCTATGCCCAAGAACAGCTGTTTGTGTAAGCATCAGTGGCCCCGACAGGGGTG 309

DB 81 AlaLeuValTyrAlaLysGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyVal 100

QY 310 CCCAAGTGTGGTGTGGTGTGACAGATGCGGCTCAGCGACCCGTGGGGCCCCCAATG 369

DB 101 ProLysValLeuValTyrValThrAspLysValSerSerAspProValGlyProPheMet 120

QY 370 CAGGAGCTCAAGAACCTGGGGGTGACCGTGTTCATTGTTCAGACCCGGCGAGCAACTTC 429

DB 121 GlnGlnLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArgGlyLysPhe 140

QY 430 CTGAGCTGTACGCCGTGCTGCTCAGCCCTGCGGAAAGCACCTGCACTTTGTGACGTG 489

DB 141 LeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspVal 160

QY 490 GATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534

DB 161 AspAspLeuHisIleIleValGlnGlnLeuValArgGlySerIleLeu 175

XX

XX RESULT 13

XX ID AAB42581

XX ID AAB42581 standard; protein; 299 AA.

XX AC AAB42581;

XX XX

XX 08-FEB-2001 (first entry)

XX DE Human ORF2345 polypeptide sequence SEQ ID NO:4690.

XX XX

XX Human; open reading frame; ORF; detection; cytosstatic; hepatotropic;

XX vulnerability; antiparasitic; antiparkinsonian; neurotoxic; neuroprotective;

XX anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;

XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

XX hypotensive; dermatological; immunosuppressive; antiinflammatory;

XX antiviral; antibacterial; antifungal; antineumatic; antithyroid;

XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;

XX neurodegenerative disorder; osteoarthritis; graft vs host disease;

CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medicament in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
CC present sequence is mouse VA domain

SQ Sequence 180 AA;

Alignment Scores:

Pred. No.:	5,31e-57	Length:	180
Score:	762.00	Matches:	149
Percent Similarity:	91.5%	Conservative:	13
Best Local Similarity:	84.2%	Mismatches:	15
Query Match:	72.6%	Indels:	0
DB:	6	Gaps:	0

US-10-699-035A-1 (1-537) x AAE32503 (1-180)

QY 1 GGGGACCTGATGTCCTGCTGACAGCTGACGAGGCTCTCTACATACAGATTCTCCCG 60
DB 2 GlyAspLeuLeuPheLeuLeuAspSerAlaSerValSerHisTyrGluPheSerArg 21
QY 61 GTTCGGAGATTGTGGGGAGCTGTGCTGCACCTGCCCTGGGACCGGGCCCTGCGT 120
DB 22 ValArgGluPheValGlyGlnLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 41
QY 121 GCCAGTCTGTGACAGCTGGGACGCTGCGCCATACACGAGTCCCTTGGCCAGCACAGC 180
DB 42 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 61
QY 181 TCGGATGAGGCTGCCAGATGCGGTGCGCTTCTGCCAGGACGATGGGTGACACCCAC 240
DB 62 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspTrpAsn 81
QY 241 ACTGCGCTGCGCTGTGCTATGCCAGAAACAGCTGTTGCTAGACATCAGGTGCCCG 300
DB 82 ThrGlyLeuAlaLeuAlaTyrAlaTyrGluGlnLeuPheAlaGlnAlaGlyAlaArg 101
QY 301 CCAGGGGCTGCCCAAGTGTGCTGTGCTGACAGATGGCGGCTCCACGACCTGTGGGC 360
DB 102 ProGlyValProGlyValLeuValIleTyrValThrAspGlyGlySerSerAspProValGly 121
QY 361 CCCCCATGAGGAGCTCAAGACCTGGGGCTGACCGGTCTATTGTGACGACCGGCCGA 420
DB 122 ProPheMetGlnGlnLeuLeuAspLeuGlyValThrIlePheIleValSerThrGlyArg 141
QY 421 GGCACCTGCTGAGCTGTGACGCGCTGCTGACCCCTGCCGAGAAGACCTGCACTTT 480
DB 142 GlyAsrLeuLeuGlnLeuLeuAlaAlaIleSerAlaProIleGlnIleHisLeuHisPhe 161
QY 481 GTTGACGTGATGACCTGACATCATTTGTCGAAGACTGAGGGGCTTCATT 531
DB 162 ValAspValAspAspLeuProIleIleIleAlaArgGlnLeuArgGlySerIle 178

RESULT 15

AAE32501

ID AAE32501 standard; protein; 415 AA.

XX AAE32501;

XX 24-MAR-2003 (first entry)

XX Mouse Willebrand Factor A domain related-protein (WARP).

XX Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
XX extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
XX gene therapy; mouse.

XX Mus sp.

PH Key Location/Qualifiers
FT Peptide 1..18
FT Protein /label=Signal_peptide
FT Protein /note=.415
FT Modified-site 148 /note="Mouse mature WARP protein"
FT Modified-site 148 /note="O-glycosylation site"
FT Modified-site 264 /note="N-glycosylation site"
FT Modified-site 359 /note="N-glycosylation site"
FT Modified-site 361 /note="N-glycosylation site"
FT Disulfide-bond 369..393 /note="O-glycosylation site"
FT Modified-site 400 /note="O-glycosylation site"
PN MO200288184-A1.
XX
XX 07-NOV-2002.
XX
XX 02-MAY-2002; 2002MO-AU000542.
XX
XX 02-MAY-2001; 2001AU-00004701.
XX
XX (MURD-) MURDOCH CHILDRENS RES INST.
XX Bateman JF, Fitzgerald DJ;
XX
XX WPI, 2003-111873/10.
XX DR N-PSDB; AAD50022, AAD50398.
XX
XX New isolated Willebrand Factor A-related protein polypeptide useful for
XX the manufacture of a medicament in the treatment of a disease condition
XX of the extracellular matrix, in particular arthritis.
XX
XX Claim 13; Page 74-75; 103pp; English.
XX
XX The invention relates to Willebrand Factor A domain related-protein
XX (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
XX superfamily of extracellular matrix (ECM) proteins. WARP is used as a
XX molecular marker, used for detecting a loss of ECM integrity in an animal
XX subject, monitoring repair, regeneration or other disease processes in an
XX animal subject and detecting a disease condition or a propensity for the
XX development of a disease condition in an animal subject. The invention is
XX useful for the manufacture of a medicament in the treatment of a disease
XX condition of the ECM. The disease condition involves the cartilage, and
XX is preferably arthritis. The invention is also used in gene therapy. The
XX present sequence is mouse WARP protein

SQ Sequence 415 AA;

Alignment Scores:

Pred. No.:	6,28e-57	Length:	415
Score:	762.00	Matches:	149
Percent Similarity:	91.5%	Conservative:	13
Best Local Similarity:	84.2%	Mismatches:	15
Query Match:	72.6%	Indels:	0
DB:	6	Gaps:	0

US-10-699-035A-1 (1-537) x AAE32501 (1-415)

QY 1 GGGGACCTGATGTCCTGCTGACAGCTGACGAGGCTCTCTACATACAGATTCTCCCG 60
DB 33 GlyAspLeuLeuPheLeuLeuAspSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCGGAGATTGTGGGGAGCTGTGCTGCACCTGCCCTGGGACCGGGCCCTGCGT 120
DB 53 ValArgGluPheValGlyGlnLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 72
QY 121 GCCAGTCTGTGACAGCTGGGACGCTGCGCCATACACGAGTCCCTTGGCCAGCACAGC 180
DB 73 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92

```
Oy 181 TCGGGTAGGCTGCCAGATGCGGTGCGTCTCTGCCAGCGCATGGGTACACCCAC 240
Db 93 SerGlyInAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 112
Oy 241 ACTGCCCTGGCGCTGTCTATATGCCAAGAACGCTGTTTGTGAAGCATCAGGTGCCCG 300
Db 113 ThrGlyLeuAlaLeuAlaTyrAlaLysGlnLeuPheAlaGlnGluAlaGlyAlaArg 132
Oy 301 CCAGGGGTGCCCAAGTGTGTGGTGAACAGATGGCGGCTCCAGCGACCCCTGTGGGC 360
Db 133 ProGlyValProlyValLeuValTrrPvalThrAspGlyGlySerSerAspProValGly 152
Oy 361 CCCCCATGCAGAGCTCAAGGACCTGGGCGTCAACGGTTCATTGTACGACCGGCGCA 420
Db 153 ProPmetGlnGlnLeuLysAspLeuGlyValThrIlePheIleValSerThrGlyArg 172
Oy 421 GGCACTTCCTGGAGCTGTCAAGCCGCTGCTCAGCCCTGCGGAGAGAAGACCTGCACCTT 480
Db 173 GlyAsnLeuLeuGlnLeuLeuAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 192
Oy 481 GTGACGTGATGATGCATCATTTGCAAGAGCTGAGGGGCTCCATT 531
Db 193 ValAspValAspAspLeuProIleIleAlaArgGlnLeuArgGlySerIle 209
```

Search completed: February 13, 2006, 13:23:30
Job time : 97.9481 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: February 13, 2006, 13:42:00 ; Search time 17.5402 Seconds
(without alignments)
2558.401 Million cell updates/sec

Title: US-10-699-035A-1

Perfect score: 1049
Sequence: 1 gggagcctgacgtctcctcgtc.....tgaagggtccatctcgcg 537

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues
Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ r2p model -DEV=xld
-O=/abse/ABSWEB_spool/US10699035/runat_13022006_062450_25572/app_query.fasta_1
-DB=Published Applications_AA_Main -OPMT=fastan -SUFFIX=rapbm -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abs02p
-USER=US10699035@cgn1.1.405@runat_13022006_062450_25572 -NCPU=6 -ICPU=3
-NO MAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	902	86.0	180	4	US-10-699-035A-2
2	902	86.0	215	3	US-09-789-561-85
3	902	86.0	215	3	US-09-833-245-2096
4	902	86.0	215	5	US-10-883-936-85
5	902	86.0	226	3	US-09-789-561-165
6	902	86.0	226	5	US-10-883-936-165
7	902	86.0	242	3	US-09-789-561-159
8	902	86.0	242	5	US-10-883-936-159
9	902	86.0	418	4	US-10-699-035A-6
10	902	86.0	418	4	US-10-699-035A-20
11	886	84.5	185	4	US-10-149-819-18

12	762	72.6	180	4	US-10-699-035A-8	Sequence 8, Appli
13	762	72.6	181	4	US-10-699-035A-31	Sequence 31, Appli
14	762	72.6	415	4	US-10-699-035A-4	Sequence 4, Appli
15	762	72.6	415	4	US-10-699-035A-21	Sequence 21, Appli
16	751	71.6	186	3	US-09-789-561-160	Sequence 160, App
17	751	71.6	186	5	US-10-883-936-160	Sequence 160, App
18	727.5	26.0	176	3	US-09-976-782-38	Sequence 38, Appli
19	265.5	25.3	182	4	US-10-699-035A-22	Sequence 22, Appli
20	263.5	25.1	184	4	US-10-699-035A-27	Sequence 27, Appli
21	261.5	24.9	3063	3	US-09-918-715-257	Sequence 257, App
22	261.5	24.9	3063	4	US-10-177-293-61	Sequence 61, Appli
23	261.5	24.9	3063	4	US-10-177-293-63	Sequence 63, Appli
24	261.5	24.9	3063	4	US-10-301-822-26	Sequence 26, Appli
25	261.5	24.9	3063	4	US-10-474-794-257	Sequence 257, App
26	261.5	24.9	3063	5	US-10-979-159-257	Sequence 257, App
27	261.5	24.9	3063	5	US-10-631-467-918	Sequence 918, App
28	261.5	24.9	3118	5	US-10-220-335-287	Sequence 287, App
29	259	24.7	580	5	US-10-722-860-2769	Sequence 2769, Ap
30	254.5	24.3	755	3	US-09-919-497-57	Sequence 57, Appli
31	254.5	24.3	1207	4	US-10-408-765A-1591	Sequence 1591, Ap
32	254.5	24.3	1297	4	US-10-187-975-102	Sequence 102, App
33	254.5	24.3	1780	4	US-10-115-479-10	Sequence 10, Appli
34	247.5	23.6	415	3	US-09-976-782-114	Sequence 114, App
35	247.5	23.6	493	3	US-09-976-782-115	Sequence 115, App
36	246.5	23.5	776	4	US-10-000-512-8	Sequence 8, Appli
37	246.5	23.5	776	4	US-10-074-566-8	Sequence 8, Appli
38	246.5	23.5	782	4	US-10-428-275-176	Sequence 176, App
39	246.5	23.5	794	3	US-09-833-245-1292	Sequence 1292, Ap
40	246.5	23.5	896	4	US-10-004-378A-16	Sequence 16, Appli
41	246.5	23.5	896	4	US-10-428-275-158	Sequence 158, App
42	246.5	23.5	914	4	US-10-428-275-156	Sequence 156, App
43	246.5	23.5	915	3	US-09-909-320-34	Sequence 34, Appli
44	246.5	23.5	915	3	US-09-909-088B-34	Sequence 34, Appli
45	246.5	23.5	915	3	US-09-905-291A-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2

Alignment Scores:
Pred. No.: 7,366-65
Score: 902.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 86.0%
DB: 4
Length: 180
Matches: 178
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-2 (1-180)
QY 1 GGGAGCTGATGTTCTCTGTCGACAGCTCAGCAGCTTCTTCACTAGAGTTCCCGG 60
|||||

Db 2 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 21
QY 61 GTTCGGAGATTGTGGGGCAGCTGTGGCTCCACTGCGCCCTGGGCGACCGGGCCCTGGGT 120
Db 22 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 41
QY 121 GCCAGCTGTGTGCAGCTGGGCGAGTCGGCCATACCGAGTTCCCTTGGGCGAGACAGC 180
Db 42 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 61
QY 181 TCGGGTGAAGCTGCCAGGATGGCGGTGCTTGGCCCGGAGATGGTGAACACCCAC 240
Db 62 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 81
QY 241 ACTGGCTGTGGCGCTGTATGTCGAAGAACAGCTGTTTGTGAAGCATCAGAGTGGCCG 300
Db 82 ThrGlyLeuAlaLeuValTyrAlaGlySerGluGlnLeuPheAlaGluAlaSerGlyAlaArg 101
QY 301 CCAGGGGTGCCCAAGTGTGTGGTGTGCAGATGGCGGCTCCAGCGACCTGTGGGC 360
Db 102 ProGlyValProGlyValLeuValTyrValThrAspGlyGlySerSerAspProValGly 121
QY 361 CCCCCATGCAGAGACTCAAGACCTGGGCGCTCACCGTGTTCATTGTGCAGACCGGCGGA 420
Db 122 ProPrometGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 141
QY 421 GGCAACTTCCTGGAGCTGTGCAGCCGCTGCAGCCCTGCCAGAGACACCTGCACCTTT 480
Db 142 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 161
QY 481 GTGACGTGATGATGATGCATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 162 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 179

RESULT 2

US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-85

Alignment Scores:

Pred. No.: 7.52e-65 Length: 215
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 3 Gaps: 0

US-10-699-035A-1 (1-537) x US-09-789-561-85 (1-215)

QY 1 GGGGACCTGANGTCTCTGCTGACAGCTCAGCCAGGCTCTCCTACCTAGAGTTTCCCGG 60

Db 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCGGAGATTGTGGGGCAGCTGTGGCTCCACTGCGCCCTGGGCGACCGGGCCCTGGGT 120
Db 53 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY 121 GCCAGCTGTGTGCAGCTGGGCGAGTCGGCCATACCGAGTTCCCTTGGGCGAGACAGC 180
Db 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGGATGGCGGTGCTTCTGCGCCGATGGTGAACACCCAC 240
Db 93 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
QY 241 ACTGGCTGTGGCGCTGTATGTCGAAGAACAGCTGTTTGTGAAGCATCAGAGTGGCCG 300
Db 113 ThrGlyLeuAlaLeuValTyrAlaGlySerGluGlnLeuPheAlaGluAlaSerGlyAlaArg 132
QY 301 CCAGGGGTGCCCAAGTGTGTGGTGTGCAGATGGCGGCTCCAGCGACCTGTGGGC 360
Db 133 ProGlyValProGlyValLeuValTyrValThrAspGlyGlySerSerAspProValGly 152
QY 361 CCCCCATGCAGAGACTCAAGACCTGGGCGCTCACCGTGTTCATTGTGCAGACCGGCGGA 420
Db 153 ProPrometGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 172
QY 421 GGCAACTTCCTGGAGCTGTGCAGCCGCTGCAGCCCTGCCAGAGACACCTGCACCTTT 480
Db 173 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 192
QY 481 GTGACGTGATGATGATGCATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210

RESULT 3

US-09-833-245-2096
; Sequence 2096, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PFS46PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2096
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

Alignment Scores:

Pred. No.: 7.52e-65 Length: 215
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 3 Gaps: 0

US-10-699-035A-1 (1-537) x US-09-833-245-2096 (1-215)

QY 1 GGGGACCTGATGTTCTCTGAGACGCTCAGCCAGCCTTCTCTACTACGAGTTCTCCCGG 60
Db 46 GlyAspLeuMet PheLeuLeuAspSerSerLaseValSerHisTyrGluPheSerArg 65
QY 61 GTTCGGGAGTTTGTGGGACAGCTGATGCTCCATGCGCCCTGGGACACCGGGGCTTGCGT 120
Db 66 ValArgGluPheValIGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 85
QY 121 GCCAGTGTGTGACAGTGGGAGTGGCCATACACGAGTTCCCTTCGGCCAGCACAGC 180
Db 86 AlaserLeuValHisValIGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 105
QY 181 TCGGGTAGGCTGCCCGAGAGTGGCGGTGCTTTGCGCCAGCGATGGGTGACACCCAC 240
Db 106 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 125
QY 241 ACTGGCTGGCGCTGTCTATAGCCCAAGGACAGCTGTTGCTGAAGCATCAGTGGCCCGG 300
Db 126 ThrGlyLeuAlaLeuValTyrAlaIbysGlnLeuPheAlaGlnHisSerGlyAlaArg 145
QY 301 CCAGGGGTGCCCAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGCAGCCTGTGGGC 360
Db 146 ProGlyValProIbysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 165
QY 361 CCCCCATGACGAGACTCAAGAGCTGGGCGCTCAGCCGTTCATTGTACAGACCGCGCGA 420
Db 166 ProPheMetGlnGluLeuIbysAspLeuGlyValThrValPheIleValSerThrGlyArg 185
QY 421 GGCAACTTCCTGGAGCTGTGACGCGCTGCTCAGCCCTGCGGAGAGACCTTGACATT 480
Db 186 GlysAnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluIbysHisLeuHisPhe 205
QY 481 GTGACGTGTGATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 206 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 223

RESULT 6

US-10-883-936-165
; Sequence 165, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 165
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-883-936-165

Alignment Scores:

Pred. No.: 7.57e-65 Length: 226
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 5 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-883-936-165 (1-226)

QY 1 GGGGACCTGATGTTCTCTGAGACGCTCAGCCAGCCTTCTCTACTACGAGTTCTCCCGG 60

Db 46 GlyAspLeuMet PheLeuLeuAspSerSerLaseValSerHisTyrGluPheSerArg 65
QY 61 GTTCGGGAGTTTGTGGGACAGCTGATGCTCCATGCGCCCTGGGACACCGGGGCTTGCGT 120
Db 66 ValArgGluPheValIGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 85
QY 121 GCCAGTGTGTGACAGTGGGAGTGGCCATACACGAGTTCCCTTCGGCCAGCACAGC 180
Db 86 AlaserLeuValHisValIGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 105
QY 181 TCGGGTAGGCTGCCCGAGAGTGGCGGTGCTTTGCGCCAGCGATGGGTGACACCCAC 240
Db 106 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 125
QY 241 ACTGGCTGGCGCTGTCTATAGCCCAAGGACAGCTGTTGCTGAAGCATCAGTGGCCCGG 300
Db 126 ThrGlyLeuAlaLeuValTyrAlaIbysGlnLeuPheAlaGlnHisSerGlyAlaArg 145
QY 301 CCAGGGGTGCCCAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGCAGCCTGTGGGC 360
Db 146 ProGlyValProIbysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 165
QY 361 CCCCCATGACGAGACTCAAGAGCTGGGCGCTCAGCCGTTCATTGTACAGACCGCGCGA 420
Db 166 ProPheMetGlnGluLeuIbysAspLeuGlyValThrValPheIleValSerThrGlyArg 185
QY 421 GGCAACTTCCTGGAGCTGTGACGCGCTGCTCAGCCCTGCGGAGAGACCTTGACATT 480
Db 186 GlysAnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluIbysHisLeuHisPhe 205
QY 481 GTGACGTGTGATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 206 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 223

RESULT 7

US-09-789-561-159
; Sequence 159, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-159

Alignment Scores:

Pred. No.: 7.63e-65 Length: 242
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 3 Gaps: 0

US-10-699-035a-1 (1-537) x US-09-789-561-159 (1-242)

QY 1 GGGAGCTGATGTTCTCTGTGAGACAGCTCAGCCAGCTCTCTACTAGAGTTCTCCGG 60
 DB 60 GYAAPLeuMet.PheLeuLeuAspSerSerAlaSerValSerHisIleTyrGluPheSerArg 79
 QY 61 GTTCCGAGATTGTGGGGAGAGCTGGGAGCTCCAGTCCCGCCCTGGGACCGGGCCCTGGCT 120
 DB 80 VALArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 99
 QY 121 GCCAGCTGTGTCAGCTGGGAGAGTCCGACATACACCGAGTTCCCTTCGGCCAGACAGC 180
 DB 100 AlAserLeuValHisIleValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 119
 QY 181 TCGGCTGAGCTGCCAGAGATGGGTGCTGCTTCTGCCAGCGCATGGGTGACACCCAC 240
 DB 120 SerGlyGluAlaIleGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 139
 QY 241 ACTGGCGTGGCGCTGCTCTATGCCAAGAGACAGCTTTGTCAGACATCAGTGGCCGG 300
 DB 140 ThrGlyLeuAlaLeuValIleTyrAlaLysGlnLeuPheAlaGlnAlaSerGlyAlaArg 159
 QY 301 CCAGGGGTGCCCAAAGTGTGCTGGTGGTGAAGATGGCGGCTCCAGCAGCCCTGGAGC 360
 DB 160 ProGlyValProLysValLeuValIleTyrValThrAspGlyGlySerSerAspProValGly 179
 QY 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCGTTTCATTGTGACACCGCCGA 420
 DB 180 ProprometGlnGluLeuLysAspLeuGlyValIleThrValPheIleValSerThrGlyArg 199
 QY 421 GGCAATCTCTGAGAGCTGTACAGCCGCTGCTCAGCCCTCCGAGAGACCTGACCTT 480
 DB 200 GlyAsnPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 219
 QY 481 GTGACGTGATGACCTGCACATCATTTGTCCAAGAGCTGAGGGGCTCCATTCTC 534
 DB 220 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 237

RESULT 8
 US-10-883-936-159
 ; Sequence 159, Application US/10883936
 ; Publication No. US20050019866A1
 ; GENERAL INFORMATION:
 ; APPLICANT: N1 et al.
 ; TITLE OF INVENTION: 52 Human secreted proteins
 ; FILE REFERENCE: P2043P1
 ; CURRENT APPLICATION NUMBER: US/10/883,936
 ; CURRENT FILING DATE: 2004-07-06
 ; PRIOR APPLICATION NUMBER: US/09/789,561
 ; PRIOR FILING DATE: 2001-02-22
 ; PRIOR APPLICATION NUMBER: PCT/US00/24008
 ; PRIOR FILING DATE: 2000-08-31
 ; PRIOR APPLICATION NUMBER: 60/152,317
 ; PRIOR FILING DATE: 1999-09-03
 ; PRIOR APPLICATION NUMBER: 60/152,315
 ; PRIOR FILING DATE: 1999-09-03
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 159
 ; LENGTH: 242
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (2)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (5)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (34)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-883-936-159

Alignment Scores:
 Pred. No.: 7.63e-65 Length: 242
 Score: 902.00 Matches: 178
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 5 Gaps: 0

US-10-699-035a-1 (1-537) x US-10-883-936-159 (1-242)

QY 1 GGGAGCTGATGTTCTCTGTGAGACAGCTCAGCCAGCTCTCTACTAGAGTTCTCCGG 60
 DB 60 GYAAPLeuMet.PheLeuLeuAspSerSerAlaSerValSerHisIleTyrGluPheSerArg 79
 QY 61 GTTCCGAGATTGTGGGGAGAGCTGGGAGCTCCAGTCCCGCCCTGGGACCGGGCCCTGGCT 120
 DB 80 VALArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 99
 QY 121 GCCAGCTGTGTCAGCTGGGAGAGTCCGACATACACCGAGTTCCCTTCGGCCAGACAGC 180
 DB 100 AlAserLeuValHisIleValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 119
 QY 181 TCGGCTGAGCTGCCAGAGATGGGTGCTGCTTCTGCCAGCGCATGGGTGACACCCAC 240
 DB 120 SerGlyGluAlaIleGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 139
 QY 241 ACTGGCGTGGCGCTGCTCTATGCCAAGAGACAGCTTTGTCAGACATCAGTGGCCGG 300
 DB 140 ThrGlyLeuAlaLeuValIleTyrAlaLysGlnLeuPheAlaGlnAlaSerGlyAlaArg 159
 QY 301 CCAGGGGTGCCCAAAGTGTGCTGGTGGTGAAGATGGCGGCTCCAGCAGCCCTGGAGC 360
 DB 160 ProGlyValProLysValLeuValIleTyrValThrAspGlyGlySerSerAspProValGly 179
 QY 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCGTTTCATTGTGACACCGCCGA 420
 DB 180 ProprometGlnGluLeuLysAspLeuGlyValIleThrValPheIleValSerThrGlyArg 199
 QY 421 GGCAATCTCTGAGAGCTGTACAGCCGCTGCTCAGCCCTCCGAGAGACCTGACCTT 480
 DB 200 GlyAsnPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 219
 QY 481 GTGACGTGATGACCTGCACATCATTTGTCCAAGAGCTGAGGGGCTCCATTCTC 534
 DB 220 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 237

RESULT 9
 US-10-699-035a-6
 ; Sequence 6, Application US/10699035A
 ; Publication No. US20040214349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bateman, John
 ; TITLE OF INVENTION: A Molecular Marker
 ; FILE REFERENCE: A36056 PCT USA A 071838.0142
 ; CURRENT APPLICATION NUMBER: US/10/699,035A
 ; CURRENT FILING DATE: 2003-10-31
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
 ; PRIOR FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: AU PRA701/01
 ; PRIOR FILING DATE: 2001-05-02
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 418
 ; TYPE: PRT


```

: ORGANISM: Homo sapiens
US-10-699-035A-6
Alignment Scores:
Pred. No.: 8,17e-65 Length: 418
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 4 Gaps: 0
US-10-699-035A-1 (1-537) x US-10-699-035A-6 (1-418)
QY 1 GGGAGCTGATGTTCTGCTGAGCAGCTCAGCCAGCCGTTCTCTCACTACGAGTTCTCCCGG 60
DB 33 G1yAspLeuMetPheLeuLeuAspSerSerAlaSerHisTyrGluPheSerArg 52
QY 61 GTTGGAGATTGTTGGGGCAGCTGGTGCACCTCCAGCCCGGAGCCGCGCTGGGT 120
DB 53 ValArgGluPheValG1yGlnLeuValAlaProLeuProLeuG1yThrG1yAlaLeuArg 72
QY 121 GCCAGCTGTGTCAGCTGGGCGAGTCGGCCATACACCGAGTTCCCTTCGGCCAGACAGC 180
DB 73 AlaSerLeuValHisValG1ySerArgProTyrThrGluPheProPheG1yGlnHisSer 92
QY 181 TCGGGGTGAGCTGCCCCAGAGATCGGGTGCCTTCTGCCAGCCGATGGGTGACACCCAC 240
DB 93 SerG1yGlnAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetG1yAspThrHis 112
QY 241 ACTGGCCTGGCGCTGCTATGCCAAGAAAGAGCTGTTTGGAGACATCAGATCCCGG 300
DB 113 ThrG1yLeuAlaLeuValTyrAlaLysGluGlnLeuPheAlaGlnAlaSerT1yAlaArg 132
QY 301 CCAGGGGTGCCCAAGTGTGCTGTGGGTGACAGATGGCGGCTCCAGCACCCTGTGGGC 360
DB 133 ProG1yValProLysValLeuValIrrValIThrAspG1yG1ySerSerAspProValG1y 152
QY 361 CCCCCATGCAGAGCTCAAGGACCTGGGCGTCAACCGTGTTCATTGTCCAGCCGCGCA 420
DB 153 ProPheMetGlnGluLeuLysAspLeuG1yValIrrValPheIleValSerThrG1yArg 172
QY 421 GGCAATTCTCTGAGAGCTGTACAGCCGCTTCAAGCCCGGAGAAAGACCTGCACATT 480
DB 173 G1yAspPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 192
QY 481 GTGAGCGTGATGACCTGCACATCTTGTCAAGAGCTGAGGGGCTCCATTCTC 534
DB 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgG1ySerIleLeu 210
RESULT 10
US-10-699-035A-20
: Sequence 20, Application US/10699035A
: Publication No. US20040214349A1
: GENERAL INFORMATION:
: APPLICANT: Bateman, John
: APPLICANT: Fitzgerald, David
: TITLE OF INVENTION: A Molecular Marker
: FILE REFERENCE: A36056 PCT USA A 071838, 0142
: CURRENT APPLICATION NUMBER: US/10/699, 035A
: CURRENT FILING DATE: 2003-10-31
: PRIOR APPLICATION NUMBER: PCT/AU02/00542
: PRIOR FILING DATE: 2002-05-02
: PRIOR APPLICATION NUMBER: AU PR4701/01
: PRIOR FILING DATE: 2001-05-02
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 20
: LENGTH: 418
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-699-035A-20
Alignment Scores:

```

```

Pred. No.: 8,17e-65 Length: 418
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 4 Gaps: 0
US-10-699-035A-1 (1-537) x US-10-699-035A-20 (1-418)
QY 1 GGGAGCTGATGTTCTGCTGAGCAGCTCAGCCAGCCGTTCTCTCACTACGAGTTCTCCCGG 60
DB 33 G1yAspLeuMetPheLeuLeuAspSerSerAlaSerHisTyrGluPheSerArg 52
QY 61 GTTGGAGATTGTTGGGGCAGCTGGTGCACCTCCAGCCCGGAGCCGCGCTGGGT 120
DB 53 ValArgGluPheValG1yGlnLeuValAlaProLeuProLeuG1yThrG1yAlaLeuArg 72
QY 121 GCCAGCTGTGTCAGCTGGGCGAGTCGGCCATACACCGAGTTCCCTTCGGCCAGACAGC 180
DB 73 AlaSerLeuValHisValG1ySerArgProTyrThrGluPheProPheG1yGlnHisSer 92
QY 181 TCGGGGTGAGCTGCCCCAGAGATCGGGTGCCTTCTGCCAGCCGATGGGTGACACCCAC 240
DB 93 SerG1yGlnAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetG1yAspThrHis 112
QY 241 ACTGGCCTGGCGCTGCTATGCCAAGAAAGAGCTGTTTGGAGACATCAGATCCCGG 300
DB 113 ThrG1yLeuAlaLeuValTyrAlaLysGluGlnLeuPheAlaGlnAlaSerT1yAlaArg 132
QY 301 CCAGGGGTGCCCAAGTGTGCTGTGGGTGACAGATGGCGGCTCCAGCACCCTGTGGGC 360
DB 133 ProG1yValProLysValLeuValIrrValIThrAspG1yG1ySerSerAspProValG1y 152
QY 361 CCCCCATGCAGAGCTCAAGGACCTGGGCGTCAACCGTGTTCATTGTCCAGCCGCGCA 420
DB 153 ProPheMetGlnGluLeuLysAspLeuG1yValIrrValPheIleValSerThrG1yArg 172
QY 421 GGCAATTCTCTGAGAGCTGTACAGCCGCTTCAAGCCCGGAGAAAGACCTGCACATT 480
DB 173 G1yAspPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 192
QY 481 GTGAGCGTGATGACCTGCACATCTTGTCAAGAGCTGAGGGGCTCCATTCTC 534
DB 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgG1ySerIleLeu 210
RESULT 11
US-10-149-819-18
: Sequence 18, Application US/10149819
: Publication No. US20030044913A1
: GENERAL INFORMATION:
: APPLICANT: INCYTE GENOMICS, INC.
: APPLICANT: YUE, Henry
: APPLICANT: AZIMZAI, Yalda
: APPLICANT: TANG, Y. Tom
: APPLICANT: PATTERSON, Chandra
: APPLICANT: BAUGHN, Mariah R.
: APPLICANT: LU, Dyrung Anna M.
: APPLICANT: SHAH, Puryi
: APPLICANT: LAU, Preeti
: APPLICANT: AU-YOUNG, Janice
: APPLICANT: BURFORD, Neil
: TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
: FILE REFERENCE: PF-0760 PCT
: CURRENT APPLICATION NUMBER: US/10/149, 819
: CURRENT FILING DATE: 2002-06-10
: PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
: PRIOR FILING DATE: 1999-12-10; 1999-12-16
: NUMBER OF SEQ ID NOS: 42
: SOFTWARE: PERL Program
: SEQ ID NO 18
: LENGTH: 185
: TYPE: PRT
: ORGANISM: Homo sapiens

```

```
FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CD1
US-10-149-819-18

Alignment Scores:
Pred. No.: 1,486-63 Length: 185
Score: 886.00 Matches: 175
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 84.5% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-149-819-18 (1-185)

QY 10 ATGTTCTGCTGAGACAGCTCAGCCAGCGCTCTCACTACAGATTCTCCGGGTTCCGGAG 69
DB 1 MetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArgValaArgGlu 20
QY 70 TTGTGGGGGACGTGTGCTGCTCACTGCCCCCTGGGACCGGGGCCCCCTGGCTGCACTGT 129
DB 21 PheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeu 40
QY 130 GTGCACGTGGGACAGTGGCCATACCCAGTTCCCTTGGCCACAGCACAGCTGGGCTGAG 189
DB 41 ValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGlu 60
QY 190 GGTGCCCAAGATGGCGGTGGCTTGTGCCCCAGCGCATGGGGTGAACCCACACAGCGCTG 249
DB 61 AlaAlaGlnAspAlaValaArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeu 80
QY 250 GCGCTGTGTATGCCAAGAACAGCTGTTGTCTGAACATCAAGTGCACCGGCCAGGGGCTG 309
DB 81 AlaLeuValTyrAlaTyrGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyVal 100
QY 310 CCCAAGTCTGTGTGTGGGTGACAGATGGCGGCTCCAGGACCCCTTGGGCCCCCAGT 369
DB 101 ProLysValLeuValaTyrValThrAspGlyGlySerSerAspProValGlyProPheMet 120
QY 370 CAGGAGCTCAAGAGACCTGGGCGCTGCACGTTTCATTCATTCAGACCGGCCGAGGACCTTC 429
DB 121 GlnGlnLeuLeuAspLeuLeuValThrValPheIleValSerThrGlyArgGlyAsnPro 140
QY 430 CTGGAAGCTGTACGCGCTGCTCAGCCCTGCCGAGAACACCTGCATTTGTGACGTG 489
DB 141 LeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLeuHisLeuHisPheValaAspVal 160
QY 490 GATGACCTGCACATCATTTGTCACAGAGCTGAGGGGCTCCATTCTC 534
DB 161 AspAspLeuHisIleIleValaGlnGlnLeuArgGlySerIleLeu 175

RESULT 12
US-10-699-035A-8
; Sequence 8, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-8
```

```
Alignment Scores:
Pred. No.: 1,836-53 Length: 180
Score: 762.00 Matches: 149
Percent Similarity: 91.5% Conservative: 13
Best Local Similarity: 84.2% Mismatches: 15
Query Match: 72.6% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-8 (1-180)

QY 1 GGGGACCGATGTTCTGCTGAGACAGCTCAGCCAGCGCTCTCACTACAGATTCTCCGGG 60
DB 2 GlyAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 21
QY 61 GTTCCGAGATTGTGTGGGACAGCTGTGCTCCACAGCCCTGGGACCGGGGCCCCCTGCGT 120
DB 22 ValaArgGlnPheValaGlyGlnLeuValaAlaThrMetSerPheGlyProGlyAlaLeuArg 41
QY 121 GCGACTGTGTGCACAGTGGGACAGTGGCCATACCCAGATTCCCTTGGGCCAGACAGC 180
DB 42 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 61
QY 181 TCGGGTGAAGCTGCGCAGATGCGGTGCTGCTTGGCCACGATGGGTGACACCCAC 240
DB 62 SerGlyGlnAlaIleArgAspAlaIleArgValaAlaProGlnArgMetGlyAspThrAsn 81
QY 241 ACTGGCCTGGGCGTGGTCTATGACCAAGAACAGCTGTTGTCTGAAGCATCAGTGGCGCGG 300
DB 82 ThrGlyLeuAlaLeuAlaTyrAlaGlyGlnGlnPheAlaGlnAlaIleValaArg 101
QY 301 CCGAGGGTGGCCCAAGTCTGTGTGAGTGAACATGAGCGGCTCCAGGACCTGTGGGCT 360
DB 102 ProGlyValProLysValLeuValaTyrValThrAspGlyGlySerSerAspProValGly 121
QY 361 CCCCCATGACAGAGACTCAAGGACCTGGGCGCTGCACCGCTTCATTCATTCAGACCGGCCGA 420
DB 122 ProprometGlnGlnLeuLeuAspLeuGlyValThrIlePheIleValSerThrGlyArg 141
QY 421 GCGAATCTTCTGAGAGCTGTACGCGCTGCTCAGCCCTGCCGAGAACACCTGCATTT 480
DB 142 GlyAsnLeuLeuGlnLeuLeuAlaAlaSerAlaProAlaGlnLeuHisLeuHisPhe 161
QY 481 GTGACGTGATGATGACCTGCACATCATTTGTCACAGAGCTGAGGGGCTCCATT 531
DB 162 ValaAspValaAspAspLeuProIleIleAlaArgGlnLeuArgGlySerIle 178

RESULT 13
US-10-699-035A-31
; Sequence 31, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: VA domain from WARP
US-10-699-035A-31

Alignment Scores:
```

Pred. No.: 1.83e-53 Length: 181
 Score: 762.00 Matches: 149
 Percent Similarity: 91.5% Conservative: 13
 Best Local Similarity: 84.2% Mismatches: 15
 Query Match: 72.6% Indels: 0
 DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-31 (1-181)

```
OY 1 GGGAGCTGATGTTCTCTGAGACAGCTCAGCCAGCGTCTCTCACTAGAGTTCTCCCG
  |||||.....|
Db 2 GlyAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 21
OY 61 GTTGGAGATTGTGGGAGAGCTGTGCTCCACTGCCCTGGGAGACCGGGCCCTGCGT 120
  |||||.....|
Db 22 ValArgGluPheValIGlyGluLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 41
OY 121 GCCAGTCTGTGACAGTGGGAGTGGCCATACACCGAGTTCCCTTGGGCGACAGCAGC 180
  |||||.....|
Db 42 AlaSerLeuValHisValIGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 61
OY 181 TCGGCTGAGAGCTGCCAGAGATGGCGTGTCTCTGCCAGCGCATGGGTAGACACCGAC 240
  |||||.....|
Db 62 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 81
OY 241 ACTGGCTGGCGCTGTCTATGTCACAGAAACAGCTGTTTGTGAGACATCAGGTGCCCGG 300
  |||||.....|
Db 82 ThrGlyLeuAlaLeuAlaTyrAlaLysGluGluPheAlaGluAlaGlyAlaArg 101
OY 301 CCAAGGAGTGGCCAAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGAGACCTCTGGGC 360
  |||||.....|
Db 102 ProGlyValProLysValLeuValIlePheValThrAspGlyGlySerSerAspProValGly 121
OY 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCGTGTATTTGATGTCAGCACCGCGCA 420
  |||||.....|
Db 122 ProPheMetGlnGluLeuLysAspLeuGlyValIleThrIlePheIleValSerThrGlyArg 141
OY 421 GGCAACTTCTGAGAGCTGTACGCGCTGCTCAGCCCTGAGCCCTCGAGAGACCTGACATT 480
  |||||.....|
Db 142 GlyAsnLeuLeuGluLeuLeuAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 161
OY 481 GTGACCTGTGATGACCTGCACATCTATTGTCCAAAGAGCTGAGGGCTCCATT 531
  |||||.....|
Db 162 ValAspValAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 178
```

RESULT 14

US-10-699-035A-4
 ; Sequence 4, Application US/10699035A
 ; Publication No. US20040214349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bateman, John
 ; APPLICANT: Fitzgerald, David
 ; TITLE OF INVENTION: A Molecular Marker
 ; FILE REFERENCE: A36056 PCT USA A 071838.0142
 ; CURRENT APPLICATION NUMBER: US/10/699,035A
 ; PRIOR FILING DATE: 2003-10-31
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
 ; PRIOR FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: AU PR4701/01
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 415
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-699-035A-4

Alignment Scores:
 Pred. No.: 2.03e-53 Length: 415
 Score: 762.00 Matches: 149
 Percent Similarity: 91.5% Conservative: 13
 Best Local Similarity: 84.2% Mismatches: 15

Query Match: 72.6% Indels: 0
 DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-4 (1-415)

```
OY 1 GGGAGCTGATGTTCTCTGAGACAGCTCAGCCAGCGTCTCTCACTAGAGTTCTCCCG 60
  |||||.....|
Db 33 GlyAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
OY 61 GTTGGAGATTGTGGGAGAGCTGTGCTCCACTGCCCTGGGAGACCGGGCCCTGCGT 120
  |||||.....|
Db 53 ValArgGluPheValIGlyGluLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 72
OY 121 GCCAGTCTGTGACAGTGGGAGTGGCCATACACCGAGTTCCCTTGGGCGACAGCAGC 180
  |||||.....|
Db 73 AlaSerLeuValHisValIGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92
OY 181 TCGGCTGAGAGCTGCCAGAGATGGCGTGTCTCTGCCAGCGCATGGGTAGACACCGAC 240
  |||||.....|
Db 93 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 112
OY 241 ACTGGCTGGCGCTGTCTATGTCACAGAAACAGCTGTTTGTGAGACATCAGGTGCCCGG 300
  |||||.....|
Db 113 ThrGlyLeuAlaLeuAlaTyrAlaLysGluGluPheAlaGluAlaGlyAlaArg 132
OY 301 CCAAGGAGTGGCCAAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGAGACCTCTGGGC 360
  |||||.....|
Db 133 ProGlyValProLysValLeuValIlePheValThrAspGlyGlySerSerAspProValGly 152
OY 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCGTGTATTTGATGTCAGCACCGCGCA 420
  |||||.....|
Db 153 ProPheMetGlnGluLeuLysAspLeuGlyValIleThrIlePheIleValSerThrGlyArg 172
OY 421 GGCAACTTCTGAGAGCTGTACGCGCTGCTCAGCCCTGAGCCCTCGAGAGACCTGACATT 480
  |||||.....|
Db 173 GlyAsnLeuLeuGluLeuLeuAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 192
OY 481 GTGACCTGTGATGACCTGCACATCTATTGTCCAAAGAGCTGAGGGCTCCATT 531
  |||||.....|
Db 193 ValAspValAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 209
```

RESULT 15

US-10-699-035A-21
 ; Sequence 21, Application US/10699035A
 ; Publication No. US20040214349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bateman, John
 ; APPLICANT: Fitzgerald, David
 ; TITLE OF INVENTION: A Molecular Marker
 ; FILE REFERENCE: A36056 PCT USA A 071838.0142
 ; CURRENT APPLICATION NUMBER: US/10/699,035A
 ; PRIOR FILING DATE: 2003-10-31
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
 ; PRIOR FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: AU PR4701/01
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 21
 ; LENGTH: 415
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-699-035A-21

Alignment Scores:
 Pred. No.: 2.03e-53 Length: 415
 Score: 762.00 Matches: 149
 Percent Similarity: 91.5% Conservative: 13
 Best Local Similarity: 84.2% Mismatches: 15
 Query Match: 72.6% Indels: 0
 DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-21 (1-415)

```
QY      1 GGGGACCTGATGTTCCCTGCTGACAGCTCAGCGGCTCTCACTACGAGTTCTCCCGG 60
Db      33 GYAAPLeuLeuPheLeuLeuSerSerAlaSerValSerHisTyrGlnPheSerArg 52
QY      61 GTTCGGAGATTGTGGGGACAGCTGGTGGCTCCACTGCCCCCTGGGACCGGGGCTTGGCT 120
Db      53 ValArgGlnPheValGlnLeuValAlaThrMetSerPheGlyProGlnAlaLeuArg 72
QY      121 GCCAGCTGGTGACAGTGGGACAGTGGGACGATACCCAGAGTCCCTTGGCCGACACAGC 180
Db      73 AlaSerLeuValHisValGlySerGlnProHisThrGlnPheThrPheAspGlnTyrSer 92
QY      181 TCGGGTGAAGCTGCCAGGATGCGGTGCTGCTTCTGCCCCAGCGCATGGGTGACACCCAC 240
Db      93 SerGlnGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 112
QY      241 ACTGCTGGGCGCTGCTATGCTCAAGGAACAGCTGTTGCTGAAGCATCAGGTGCCCGG 300
Db      113 ThrGlnLeuAlaLeuAlaTyrAlaTyrGlnGlnLeuPheAlaGlnAlaGlyAlaArg 132
QY      301 CCAGGGGTGCCCAAGTGTGTGGTGGGTGACAGATGGCGCTCCAGCGACCCCTGTGGGC 360
Db      133 ProGlnValProLysValLeuValIlePvalThrAspGlyGlySerSerAspProValGly 152
QY      361 CCCCCATGCAAGAGCTCAAGAGCTGGGCGTCAACCGTTCATTGTCAGACCGGCGCA 420
Db      153 ProPheMetGlnGlnLeuLysAspLeuGlnValThrIlePheIleValSerThrGlyArg 172
QY      421 GGCAACTTCCTGAGAGCTGTCAAGCGCTGCTCAAGCCCTGCGAGAGACACTGCACATT 480
Db      173 GlnAsnLeuLeuGlnLeuLeuAlaIleAlaSerAlaProAlaGlnLysHisLeuHisPhe 192
QY      481 GTGACGTGGATGACCTGCATCATTCATTCACAAAGCTGAGGGGCTCCATT 531
Db      193 ValAspValAspAspLeuProIleIleAlaArgGlnLeuArgGlySerIle 209
```

Search completed: February 13, 2006, 13:53:23
Job time : 91.701 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:39:36 / Search time 3.95779 Seconds
(without alignments)
2243.517 Million cell updates/sec

Title: US-10-699-035A-1

Perfect score: 1049

Sequence: 1 ggggagctcgtatctctcgtc.....tgaagggtccatctcgcgcg 537

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p model -DEV=xlp
-O=/abs/ABSWEB_spool/US10699035/runat_13022006_062447_25513/app_query.fasta_1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-HOST=abbes02p -USER=US10699035 @CGN 1.1 101 @runat_13022006_062447_25513
-NCPU=6 -ICPU=3 -NO MAP -NEG SCORES=0 -WAIT -DSEPLock=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued Patents AA:*
2: /cgn2_6/prodata/1/1aa/5 COMB pep:*
3: /cgn2_6/prodata/1/1aa/6 COMB pep:*
4: /cgn2_6/prodata/1/1aa/H COMB pep:*
5: /cgn2_6/prodata/1/1aa/RE COMB pep:*
6: /cgn2_6/prodata/1/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384	36.6	77	2	US-09-513-999C-6277 Sequence 6277, Ap
2	259	24.7	584	2	US-09-949-016-10340 Sequence 10340, A
3	259	24.7	584	2	US-09-949-016-10341 Sequence 10341, A
4	254.5	24.3	755	2	US-09-919-497-57 Sequence 57, Appl
5	246.5	23.5	638	1	US-08-897-443-1 Sequence 1, Appl
6	246.5	23.5	915	2	US-09-907-794A-34 Sequence 34, Appl
7	246.5	23.5	915	2	US-09-905-125A-34 Sequence 34, Appl
8	246.5	23.5	915	2	US-09-902-775A-34 Sequence 34, Appl
9	246.5	23.5	915	2	US-09-906-700-34 Sequence 34, Appl
10	246.5	23.5	915	2	US-09-903-603A-34 Sequence 34, Appl
11	246.5	23.5	915	2	US-09-904-920A-34 Sequence 34, Appl
12	246.5	23.5	915	2	US-09-909-064-34 Sequence 34, Appl

13	246.5	23.5	915	2	US-09-905-381A-34 Sequence 34, Appl
14	246.5	23.5	915	2	US-09-906-618-34 Sequence 34, Appl
15	246.5	23.5	915	2	US-09-906-646-34 Sequence 34, Appl
16	246.5	23.5	915	2	US-09-904-462-34 Sequence 34, Appl
17	246.5	23.5	915	2	US-09-902-735A-34 Sequence 34, Appl
18	246.5	23.5	915	2	US-09-906-722A-34 Sequence 34, Appl
19	246.5	23.5	956	2	US-09-949-016-6215 Sequence 6215, Ap
20	246.5	23.5	963	2	US-09-949-016-11519 Sequence 11519, A
21	246.5	23.5	963	2	US-09-949-016-11520 Sequence 11520, A
22	237.5	22.6	956	1	US-08-897-443-3 Sequence 37, Appl
23	233.5	22.3	496	1	US-08-462-128-37 Sequence 37, Appl
24	233.5	22.3	496	1	US-08-463-180-37 Sequence 37, Appl
25	233.5	22.3	496	1	US-08-001-078A-1 Sequence 1, Appl
26	233.5	22.3	496	1	US-08-897-443-4 Sequence 4, Appl
27	233.5	22.3	496	1	US-08-463-218-1 Sequence 1, Appl
28	233.5	22.3	496	2	US-09-949-016-11306 Sequence 11306, A
29	233.5	22.3	496	4	PCP-US94-00253-1 Sequence 1, Appl
30	231.5	22.1	452	2	US-09-914-259-34 Sequence 34, Appl
31	227	21.6	405	2	US-09-912-285C-374 Sequence 374, App
32	224.5	21.4	481	2	US-09-914-259-35 Sequence 35, Appl
33	214.5	20.4	486	2	US-09-914-259-36 Sequence 36, Appl
34	214.5	20.4	486	2	US-09-976-594-278 Sequence 35, Appl
35	214.5	20.4	486	2	US-09-949-016-6216 Sequence 6216, Ap
36	214.5	20.4	507	2	US-09-949-016-9678 Sequence 9678, Ap
37	198.5	18.9	171	2	US-09-996-611D-2 Sequence 2, Appl
38	198.5	18.9	954	2	US-09-996-611D-1 Sequence 1, Appl
39	197	18.8	214	2	US-10-061-658-6 Sequence 6, Appl
40	197	18.8	214	2	US-10-061-658-9 Sequence 9, Appl
41	192	18.3	550	2	US-09-907-794A-227 Sequence 227, App
42	192	18.3	550	2	US-09-905-125A-227 Sequence 227, App
43	192	18.3	550	2	US-09-902-775A-227 Sequence 227, App
44	192	18.3	550	2	US-09-906-700-227 Sequence 227, App
45	192	18.3	550	2	US-09-579-288-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-513-999C-6277
Sequence 6277, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Maline Edwards, J.B.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59. US2, REG
CURRENT APPLICATION NUMBER: US/09/513, 999C
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122, 487
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 6277
LENGTH: 77
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: 66
OTHER INFORMATION: Xaa=Cys or Gly
US-09-513-999C-6277

Alignment Scores:

Pred. No.: 2.51e-26
Score: 384.00
Percent Similarity: 98.7%
Best Local Similarity: 98.7%
Query Match: 36.6%
DB: 2
Length: 77
Matches: 76
Conservative: 0
Mismatch: 1
Indels: 0
Gaps: 0

US-10-699-035A-1 (1-537) x US-09-513-999C-6277 (1-77)

```
QY 226 ATGGGTGACACCCACACTGGCGCTGGCTGTATGSCAAGAAACAGCTGTTTCTGAA 285
Db 1 MetGlyAspThrHisThrGlyLeuAlaLeuValTyrAlaLysGlnLeuPheAlaGln 20
QY 286 GCATCAGGTGCCCCCGCCAGGGGTGCCCAAGTCTGTGTGGGTACAGATGGCGGCTCC 345
Db 21 AlaSerGlyAlaArgProGlyValProLysValLeuValTyrValThrAspGlyGlySer 40
QY 346 AGCGACCTGTGGGCCCCCCCATGACAGAGCTCAAGACCTGGGCGGTACCGCTTCATT 405
Db 41 SerAspProValGlyProPheMetGlnGluLeuLysAspLeuGlyValThrValPheIle 60
QY 406 GTCAAGACCGGCGGAGCACTTCTGTGAGCTGTGACCGCTGCTCAGCC 456
Db 61 ValSerThrGlyArg**AsnPheLeuGluLeuSerAlaAlaAlaSerAla 77

RESULT 2
US-09-949-016-10340
; Sequence 10340, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10340
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10340

Alignment Scores:
Pred. No.: 6,256-15 Length: 584
Score: 259.00 Matches: 67
Percent Similarity: 51.1% Conservative: 25
Best Local Similarity: 37.2% Mismatches: 84
Query Match: 24.7% Indels: 4
DB: 2 Gaps: 2

US-10-699-035a-1 (1-537) x US-09-949-016-10340 (1-584)
QY 4 GACCTGATGTTCTGCTGAGACGCTCAGCGCTCTCACTAGAGTTCTCCGGGTT 63
Db 348 AspLeuValLeuLeuValAspGlySerLysSerValArgProGlnAsnPheGluLeuVal 367
QY 64 CGGAGATTGTGGGCGACCTGTGCTCCTCAGCTGCCCTGGGACCCGGGCGCTGCTGCC 123
Db 368 LysArgPheValAsnGlnIleValAspPheLeuAspValSerProGlnGlyThrArgVal 387
QY 124 AGTCTGTGACAGTGGGACGTGGCCATACACCGAGTTCCCTTGGCGGACAGACGCTCG 183
Db 388 GlyLeuValGlnPheSerSerArgValArgThrGlnPheProLeuGlnYArgTyrGlyThr 407
QY 184 GGTAGGCTGCCAGATGCGGTGCGTGTCTTGTGCCCGATGGGTGACACCCACACT 243
Db 408 AlaAlaGluValLysGlnAlaValLeuAlaValGluTyrMetGlnArgGlyThrMetThr 427
QY 244 GGCCTGGCTGTGCTATGCCAAGAACAGCTGTTGCTGAAGCATCAGTGGCCGGGCA 303
Db 428 GlyLeuAlaLeuArgHisMetValGlnHisSerPheSerGlnAlaGlnGlyAlaArgPro 447
QY 304 -----GGGCTGCCCAAGTGTGCTGTGGGTGACAGATGGCGGCTCCAGCACCT 354
```

```
Db 448 ArgAlaLeuAsnValProArgValGlyLeuValPheThrAspGlyArgSerGlnAspAsp 467
QY 355 GTGGGCCCCCATGACAGACCTCAGGACCTGGGCGGCGGAGTTCATTGTCAGCACCC 414
Db 468 HisSerValTyrPalaAlaArgAlaLysGlnGluGlyIleValMetTyrAlaValGlyVal 487
QY 415 GCGCGAGGCAACTTCTGAGACTGTACGCCCTGCTCAGCCCTGACCCGAGAACGACCTG 474
Db 488 GlyLysAlaValAlaGluAlaGluLeuArgGluIleLeuAlaSerGluProAlaGluLeuHisVal 507
QY 475 CACTTTG---GACGTGGATGACCTGCACATCTATTGTCACAGCTGAGGGGCTCCATT 531
Db 508 SerTyrAlaProAspPheGlyThrMetThrHisLeuGluAsnLeuArgGlySerIle 527

RESULT 3
US-09-949-016-10341
; Sequence 10341, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10341
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10341

Alignment Scores:
Pred. No.: 6,256-15 Length: 584
Score: 259.00 Matches: 67
Percent Similarity: 51.1% Conservative: 25
Best Local Similarity: 37.2% Mismatches: 84
Query Match: 24.7% Indels: 4
DB: 2 Gaps: 2

US-10-699-035a-1 (1-537) x US-09-949-016-10341 (1-584)
QY 4 GACCTGATGTTCTGCTGAGACGCTCAGCGCTCTCACTAGAGTTCTCCGGGTT 63
Db 348 AspLeuValLeuLeuValAspGlySerLysSerValArgProGlnAsnPheGluLeuVal 367
QY 64 CGGAGATTGTGGGCGACCTGTGCTCCTCAGCTGCCCTGGGACCCGGGCGCTGCTGCC 123
Db 368 LysArgPheValAsnGlnIleValAspPheLeuAspValSerProGlnGlyThrArgVal 387
QY 124 AGTCTGTGACAGTGGGACGTGGCCATACACCGAGTTCCCTTGGCGGACAGACGCTCG 183
Db 388 GlyLeuValGlnPheSerSerArgValArgThrGlnPheProLeuGlnYArgTyrGlyThr 407
QY 184 GGTAGGCTGCCAGATGCGGTGCGTGTCTTGTGCCCGATGGGTGACACCCACACT 243
Db 408 AlaAlaGluValLysGlnAlaValLeuAlaValGluTyrMetGlnArgGlyThrMetThr 427
QY 244 GGCCTGGCTGTGCTATGCCAAGAACAGCTGTTGCTGAAGCATCAGTGGCCGGGCA 303
Db 428 GlyLeuAlaLeuArgHisMetValGlnHisSerPheSerGlnAlaGlnGlyAlaArgPro 447
QY 304 -----GGGCTGCCCAAGTGTGCTGTGGGTGACAGATGGCGGCTCCAGCACCT 354
Db 448 ArgAlaLeuAsnValProArgValGlyLeuValPheThrAspGlyArgSerGlnAspAsp 467
```



```

Db      97 GlyLeuLeuGlnTrrTyrSerThrValValyAsnGluPheSerLeuYsthrPheIlybAyg 116
      |||::: ||||| |||||
Oy      184 GGTGAGGCTGCCAGGAGTCCGGTGCCTGCTTCCACAGCGCATGGGTGACACCCACACT 243
      :: |||::: |||
Db      117 LysSerGIuValIGluArgAlaValIlybArgMetArgH;IleuSerthrGIuThrMetThr 136
      244 GGCCTGGCCGCTGCTCATATCCCAAGAAACAAGCTTTTGTCTGAAACATCAGGTGCCCGGCA 303
      |||::: ||||| |||
Db      137 GlyLeuAlaIleGIuTrpAlaLeuAsnIleAlaPheSerGIuAlaGluGlyAlaArgPro 156
      304 -----GGGGTGGCCAAAGCTGCTGTGGGTGCACAGATGGCGGCTCCAGCACCT 354
      |||::: ||||| |||
Db      157 LeuArgGIuAsnValProArgValIleMetIleValThrAspIlyArgProGlnAspSer 176
      177 ValAlaGluValAlaAlaIlybAlaArgAspThrGIyIleuIlePheAlaIleGIyVal 196
      |||::: |||::: |||
Oy      355 GTGGGCCCCCATGACAGAGCTCAAGACCTGGGGCGTCACCGTTGATTTGTACAGACC 414
      |||::: |||::: |||
Db      415 GGGCCAGGCACTCTCTGGAGCTGTACGCCGCTCTACGCCCTTGGCCGGAAGACACTG 474
      |||::: |||::: |||
Db      197 GlyGIuValAspPheAsnThrLeuYsSerIleGIySerGIuProHIsGluAspHIsVal 216
      475 -----CACTTGTGGACGCGATGACCTGCACATCATTTGCCAAGAG 516
      |||::: |||::: |||
Db      217 PheLeuValAlaAsnPheSerGIuIleGIuThrLeuThrSerValPheGlnIlys 234

RESULT 6
US-09-907-794A-34
; Sequence 34, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907, 794A
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143, 048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145, 698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146, 222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944

```

```

PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-794A-34

Alignment Scores:
Pred. No.: 9,18e-14 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x US-09-907-794A-34 (1-915)
OY 4 GACCTGATGTCCTGCTGGACAGCTCACCCAGCGCTCTCACTACGAGATTCTCCCGGGTT 63
Db 57 AspleuValpHelletleAsperSertergSerValaBenthriSABRYrAlaIyVal 76
OY 64 CGGAGATTGTGGGCAAGCTGTGGCTCCACTGCCCCCTGGGACCGGGGCCCTGGTCC 123
Db 77 LysGluPheailleValAsprilleuGlnPheleuAaprllegIyProAerValThrArgVal 96
OY 124 AGCTCTGTCACAGCTGGGAGCTGGGCGCATACACCGAGTTCCCTTCGGCCAGACAGCTCG 183
Db 97 GlyLeuIleuGlnIlyrIdlySerTherValIlySaBngIupPheSerIleuIySthrPheIyArg 116
OY 184 GGTGAGGCTGCCAGGATGCGGTGGCTCTTCTGCCGAGCGATGGGTGACACCCACT 243
Db 117 LysSerGluValaIgluArgAlaValIlySaBngIyMetArghIlySeuSerThngIlyThrMetThr 136
OY 244 GGCCTGCGCTGGTCTATGCTCCAAAGAACAGCTGTTTGCTGAAGCATGAGTGGCCGGCA 303
Db 137 GlyLeuAlaIleGlnIlyrAlaIleuBsnIleAlaPheSerGluValaIgluIlyAlaArgPro 156
OY 304 -----GGGGTCCCAAAAGTCTGGTGGGTGGGTGAACAGTGGCGGCTCCAGGACCT 354
Db 157 LeuArgGluAaBnValProArGValIleMetIleValThrAerGlyArGProGlnAaSPSer 176
OY 355 GTGGGCGCCCCCATGAGAGAGCTCAAGAGCTGGGCGTCAACGCGTTCATTGTGACGACC 414
Db 177 ValAlaGluValaIleAlaIlySaIaArgAapThngIlyIleuIlePheAlaIleGlyVal 196
OY 415 GGCCTGAGCAACTTCTCGAGCTGTCAGCCGCTGCTCAGCCGCTCCGAGAAAGCACTG 474
Db 197 GlyGlnAlaAspPheBsnThrIleuIlySerIleGlySerGluPurrohIlyBgluAerPheIleVal 216
OY 475 -----CACTTTGTGACGTGGATGACCTGACATCATATTGTCAAAG 516

```



```
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-603A-34

Alignment Scores:
Pred. No.: 9,18e-14      Length: 915
Score: 246.50           Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5%      Indels: 7
DB: 2                   Gaps: 2

US-10-699-035A-1 (1-537) x US-09-903-603A-34 (1-915)

QY 4 GACCTGATGTTCTCTGAGACGTCAGCCAGCTCTCACTAGAGTTCTCCGGGTT 63
Db 57 AspleuVal1Phe1le1leAspSerSerArgSerValAsnThrHisAspTyrAlaVal 76
QY 64 CGGAGTTTGAGGGACAGCTGGTGGCTCCACTGCTGGGACCGGGCCCTGGTGGC 123
Db 77 LysGluPhe1leValAsp1leuGlnPhe1leuAsp1leuGlyProAspValThrArgVal 96
QY 124 AGCTGTGTCAGCTGGGAGCTGGCCATACACCGAGTTCCCTTGGCCAGACAGCTCG 183
Db 97 Gly1eu1euGlnTyrGlySerThrValLysAsnGluPheSer1euLysThrPheLysArg 116
QY 184 GGTAGGCTCCCGACAGATGCGGTGGCTGTTCGCCACGGATGGGTGACACCCACT 243
Db 117 LysSerGluValAlaArgAlaValAlaLysArgMetArgHis1euSerThrGlyThrmethr 136
QY 244 GACCTGAGCTGTGTATGTCAGGAGACAGCTGTTGCTGAAGCATCAGTCCCGGCA 303
Db 137 Gly1euAla1leGlnTyrAla1euAsn1leAlaPheSerGluAlaGluGlyAlaArgPro 156
QY 304 -----GGGGTCCCAAGTGTGTGTGGTGGTGCAGATGGCGGCTCCAGCGACT 354
Db 157 LeuArgGluAsnValProArgVal1leMet1leValThrAspGlyArgProGlnAspSer 176
QY 355 GTGGCCCCCCCATGACAGAGCTCAAGACCTGGGCGCTACCGGTTCATTGTCCAGCAC 414
Db 177 ValAlaGluValAlaAlaLysAlaArgAspThrGly1le1lePheAla1leGlyVal 196
QY 415 GCGCCAGGACCACTTCTGTGAGCTGTGAGCCGCTGCTCAGCCCTGCGGAGAGCACTGG 474
Db 197 Gly1euValAspPheAsnThr1euLysSer1leGlySerGluProHisGluAspHisVal 216
QY 475 -----CACTTGTGAGCTGTGATGACCTGCACATCATTTGTCACAGAG 516
Db 217 Phe1euValAlaAsnPheSerGln1leGluThrLeuThrSerValPheGlnLys 234

RESULT 11
US-09-904-920A-34
; Sequence 34, Application US/09904920A
; Patent No. 6806352
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerltzen, Mary E.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
```


Qy 475 -----CACTTGTGACGTGATGACCTGCACATCATTTGCCAAG 516
Db 217 PheUValAlaAaPhSeRcGlnIleGluThrLeuThrSerValPheGlnIlys 234

RESULT 13
US-09-905-381A-34
; Sequence 34, Application US/09905381A
; Patent No. 6818746
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Guiney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,381A
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219

; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-381A-34

Alignment Scores:
Pred. No.: 9,18e-14 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Gaps: 7
DB: 2

US-10-699-035A-1 (1-537) x US-09-905-381A-34 (1-915)

Qy 4 GACCTGATGTTCTGCTGACAGCTCAGCCAGCGTCTCACTACAGATTCTCCGGGTT 63
Db 57 AspleuValPheIleIleAspSerSerArgSerValAanThrHisAspTyrAlaIysVal 76

Qy 64 CCGAGATTGTGGGACAGCTGTCCTCCATCGCCCTGGGACCGGGGCGCTGGCTGCC 123
Db 77 LysGluPheIleValAspIleuGlnPheLeuAspIleGlyProAspValThrArgVal 96

Qy 124 AGCTGTGACAGCTGGGACGTCGCCATPACACGAGTTCCTTCGGCCACAGACTCG 183
Db 97 GlyLeuLeuGlnTyrGlySerThrValLysAenGluPheSerLeuLysThrPheLysArg 116

Qy 184 GGTGAGCTGCCACGATGCCGTGCTCTTCTGCCACGCCATGGGTGACACCACT 243
Db 117 LysSerGluValGluArgAlaValLysArgMetArgHisLysSerThrGlyThrMetThr 136

Qy 244 GGCTGGCGCTGCTGCTATGCAAGAAACAGCTGTTGTGAAGCATGAGCCGGCA 303
Db 137 GlyLeuAlaIleGlnTyrAlaLeuAenIleAlaPheSerGluAlaGluAlaArgPro 156

Qy 304 -----GGGTCGCCAAGTGTGTGTGTGTGACAGATGGCGCTCCAGCACT 354
Db 157 LeuArgGluAenValProArgValIleMetIleValThrArgGlyArgProGlnAspSer 176

Qy 355 GTGGGCCCCCAGACGAGCTCAAGACCTGCGCGCTACCGCTTATTTGTACGACC 414
Db 177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196

Qy 415 GGCGGAGCACTTCTGAGCTGTGACGCGCTGCTGAGCCCGCCGAGAGACACTG 474
Db 197 GlyIleValAspPheAenThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216

Qy 475 -----CACTTGTGACGTGATGACCTGCACATCATTTGCCAAG 516
Db 217 PheUValAlaAaPhSeRcGlnIleGluThrLeuThrSerValPheGlnIlys 234

RESULT 14
US-09-906-618-34
; Sequence 34, Application US/09906618
; Patent No. 6828146
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.


```

1 PRIOR APPLICATION NUMBER: PCT/US99/20554
2 PRIOR FILING DATE: 1999-09-08
3 PRIOR APPLICATION NUMBER: PCT/US99/20944
4 PRIOR FILING DATE: 1999-09-13
5 PRIOR APPLICATION NUMBER: PCT/US99/21090
6 PRIOR FILING DATE: 1999-09-15
7 PRIOR APPLICATION NUMBER: PCT/US99/21547
8 PRIOR FILING DATE: 1999-09-15
9 PRIOR APPLICATION NUMBER: PCT/US99/23089
10 PRIOR FILING DATE: 1999-10-05
11 PRIOR APPLICATION NUMBER: PCT/US99/28214
12 PRIOR FILING DATE: 1999-11-29
13 PRIOR APPLICATION NUMBER: PCT/US99/28313
14 PRIOR FILING DATE: 1999-11-30
15 PRIOR APPLICATION NUMBER: PCT/US99/28564
16 PRIOR FILING DATE: 1999-12-02
17 PRIOR APPLICATION NUMBER: PCT/US99/28565
18 PRIOR FILING DATE: 1999-12-02
19 PRIOR APPLICATION NUMBER: PCT/US99/30095
20 PRIOR FILING DATE: 1999-12-16
21 PRIOR APPLICATION NUMBER: PCT/US99/30911
22 PRIOR FILING DATE: 1999-12-20
23 PRIOR APPLICATION NUMBER: PCT/US99/30999
24 PRIOR FILING DATE: 1999-12-20
25 PRIOR APPLICATION NUMBER: PCT/US00/00219
26 PRIOR FILING DATE: 2000-01-05
27 NUMBER OF SEQ ID NOS: 423
28 SEQ ID NO 34
29 LENGTH: 915
30 TYPE: PRT
31 ORGANISM: Homo sapiens
32 IS-09-906-646-34

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:23:58 ; Search time 10.0124 Seconds
(without alignments)
2410.129 Million cell updates/sec

Title: US-10-699-035A-5

Perfect score: 2380

Sequence: 1 atgcctccctggagcgagcgcct.....ccgcacgcctggagcgctaa 1254

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p model -DEV=rlp
-Q=/abs/ABSSWEB.spool/US10699035/runat.13022006.062444.25463/app.query.fasta_1
-DB=PIR -QFMT=fastan -SUFPTX=tpir -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs802p
-USER=US10699035 @CN 1.1.77 @runat.13022006.062444.25463 -NCPU=6 -ICU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSRBLCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YFAGAP=6 -FAGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: PIR 80:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	444	18.7	3124	2 A40020	collagen alpha 1(X
2	387	16.3	2944	2 A54849	collagen alpha 1(V
3	382	16.1	1857	2 S31212	collagen alpha 1(X
4	382	16.1	1888	2 S78476	collagen alpha 1(X
5	371	15.6	1747	2 A45974	collagen alpha 1(X
6	272.5	11.4	929	2 I51027	type XII collagen
7	268	11.3	843	2 A40970	undulin 1 - human
8	263	11.1	660	1 Q00B3	BHLFI protein - hu
9	248.5	10.4	493	2 A33809	cartilage matrix p
10	246.5	10.4	1106	2 J00405	cartilage matrix p
11	243.5	10.2	500	2 S66522	cartilage matrix p
12	243.5	10.2	660	1 Q00B3	BHLFI protein - hu
13	240	10.1	1344	1 A35175	mucin 1 precursor,
14	233.5	9.8	496	2 A37979	cartilage matrix p

15	230.5	9.7	3137	2 A37797	collagen alpha 3(V
16	227	9.5	3176	2 CGH03A	collagen alpha 3(V
17	226	9.5	724	2 A48569	antigen Em100 - E1
18	221.5	9.3	460	2 T33110	hypothetical prote
19	221	9.3	1446	1 A45344	immediate-early pr
20	219	9.2	1042	1 CGCH1S	collagen alpha 1(I)
21	218.5	9.2	1460	1 EDBE1F	immediate-early pr
22	218	9.2	1042	1 CGCH1S	collagen alpha 1(I)
23	217.5	9.1	779	1 CGB01S	collagen alpha 1(I)
24	217	9.1	1027	2 S28774	collagen alpha cha
25	216.5	9.1	1453	2 S21626	collagen alpha 1(I)
26	216	9.1	1151	2 T18535	high molecular mas
27	216	9.1	1173	1 A43291	collagen alpha 2(I)
28	216	9.1	1497	2 I49607	procollagen type V
29	215	9.0	1496	1 CGH02V	collagen alpha 2(V
30	214	9.0	964	1 CGCH2S	collagen alpha 2(I)
31	213.5	9.0	712	2 A45638	immunodominant mic
32	213	8.9	741	2 T46488	hypothetical prote
33	212.5	8.9	1151	2 A45226	integrin alpha-1 c
34	212.5	8.9	1366	1 CGH02S	collagen alpha 2(I)
35	212	8.9	886	2 I50694	collagen alpha 1(I)
36	212	8.9	1464	1 CGH01S	collagen alpha 1(I)
37	211	8.9	1464	1 CGH01S	collagen alpha 1(I)
38	210.5	8.8	1466	1 CGH07L	collagen alpha 1(I)
39	210	8.8	1453	2 S21626	collagen alpha 1(I)
40	209.5	8.8	1414	1 S23809	collagen alpha 2(I)
41	209	8.8	1414	1 S23809	collagen alpha 2(I)
42	208.5	8.8	779	1 CGB01S	collagen alpha 1(I)
43	208.5	8.8	1460	1 EDBE1F	immediate-early pr
44	207.5	8.7	1446	1 A45344	immediate-early pr
45	207.5	8.7	1958	2 B40505	hypothetical prote

ALIGNMENTS

RESULT 1
A40020
collagen alpha 1(XII) chain precursor - chicken
N:Alternate names: fibrochimerin
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #ext change 09-Jul-2004
C:Accession: A40020; A34485; B34485; A28037; S23814; S22554; S28811
J:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Oba
R.; Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
U. Biol. Chem. 264, 19772-19778, 1989
A:Title: Type XII collagen. A large multidomain molecule with partial homology to type
A:Reference number: A34485; MUID:90062079; PMID:2584192
A:Accession: A34485
A:Molecule type: mRNA
A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A:Cross-references: UNIPARC:UPI0000171233; EMBL:U05137; NID:G211284; PIDN:AAA48635.1; P
A:Accession: B34485
A:Molecule type: protein
A:Residues: 2772-2792; 2846-2873 <GOR2>
A:Cross-references: UNIPARC:UPI0000173C3F; UNIPARC:UPI0000173C40
R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA
A:Reference number: A28037; MUID:87317590; PMID:3476925
A:Accession: A28037
A:Molecule type: mRNA
A:Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
A:Cross-references: UNIPARC:UPI00001712P4; EMBL:M17375; NID:G211649; PIDN:AAA48718.1; P
A:Note: this sequence has been revised in reference A34485

R:Koch, M.; Bernasconi, C.; Chiquet, M.
A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A:Reference number: S23814; PMID:92362621; PMID:1323460
A:Accession: S23814
A:Molecule type: protein
A:Residues: 'X',1333,'Q',1335-1347,1914-1928,2504,'X',2506,'X',2508-2511,'X',2513-2517
A:Cross-references: UNIPARC:PI0000173C41; UNIPARC:PI0000173C42; UNIPARC:PI0000173C43
R:Hubler, B.; van der Rest, M.
J: Biol. Chem. 262, 17724-17727, 1987
A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-
A:Reference number: S22254; PMID:88087065; PMID:3121603
A:Accession: S22254
A:Molecule type: protein
A:Residues: 2831-2832, 'T',2834,'R',2836-2843,3002-3014 <DNB>
A:Cross-references: UNIPARC:PI0000173C44; UNIPARC:PI0000173C45
R:Trzeb, J.; Trzeb, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A:Title: The two splice variants of collagen XII share a common 5' end.
A:Reference number: S28811; PMID:93042014; PMID:1420368
A:Accession: S28811
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24,1189-1257,'S',1259-1263,'E',1265-1280 <TRU>
A:Cross-references: UNIPARC:PI0000173C46; EMBL:X67327
A:Genetics:
A:Insertions: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
A:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfide
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XI) chain #status predicted <MAT>
F:24,1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted
F:24-114/Domain: IIIA #status predicted <IIIA>
F:24-105/Domain: fibronectin type III repeat homology <FN3A>
F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>
F:333-425/Domain: IIB #status predicted <IIB>
F:333-414/Domain: fibronectin type III repeat homology <FN3B>
F:461-601/Domain: von Willebrand factor type A repeat homology <VWA2>
F:62-1178/Domain: IIC #status predicted <IIC>
F:630-711/Domain: fibronectin type III repeat homology <FN3C>
F:721-802/Domain: fibronectin type III repeat homology <FN3D>
F:812-895/Domain: fibronectin type III repeat homology <FN3E>
F:905-986/Domain: fibronectin type III repeat homology <FN3F>
F:1095-1076/Domain: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>
F:1384-2295/Domain: IIID #status predicted <IIID>
F:1384-1465/Domain: fibronectin type III repeat homology <FN3I>
F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>
F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>
F:1565-1738/Domain: fibronectin type III repeat homology <FN3L>
F:1766-1838/Domain: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>
F:2119-2199/Domain: fibronectin type III repeat homology <FN3Q>
F:2207-2294/Domain: fibronectin type III repeat homology <FN3R>
F:2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>
F:2438-2440/Region: cell adhesion #status predicted
F:2509-2750/Domain: IIV, homologous to NC4 domain of type IX collagen #status predicted
F:2751-2902/Domain: collagenous COL2 #status predicted <COL2>
F:2899-2901/Region: cell attachment (R-G-D) motif
F:2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>
F:2946-3048/Domain: collagenous COL1 #status predicted <COL1>
F:3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>
F:1032,1006,1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (Asn) (cov
F:2780,2789,2836,2842,2860,2866,2869,3004,3007/Modified site: hydroxyproline (Pro) #statu

Alignment Scores:	
Pred. No.:	2.97e-18
Score:	444.00
Percent Similarity:	47.2%
Best Local Similarity:	31.1%
Query Match:	18.7%
Length:	3124
Matches:	123
Conservative:	64
Mismatches:	169
Indels:	40

DB:	2	7	Gaps:
DB:	US-10-699-035A-5 (1-1254) x A40020 (1-3124)		
QY	CGAGGAGACATGATGTTCTCTGAGACAGCTCAGCCAGCTCTCTCACTACGATTTCTCC	153	
Db	94 CGAGGAGACATGATGTTCTCTGAGACAGCTCAGCCAGCTCTCTCACTACGATTTCTCC	153	
	437 LysAlaAspValValPheLeuValAspGlySerTyrSerIleGlyIleAlaSerVal	456	
QY	CGGGTTGGAGTTGTTGGGGCAGCTGTGTGCTCCATGCCCCGTGGCACCAGGGCCCTG	213	
Db	154 CGGGTTGGAGTTGTTGGGGCAGCTGTGTGCTCCATGCCCCGTGGCACCAGGGCCCTG	213	
	457 LysValAlaGAlaPheLeuGluValLeuValLysSerPheGluIleSerProArgLysVal	476	
QY	CGTGCACATCTCTGCTCAGCTGGGACAGTGGCCATACACCGAATTCCTCTTCCGACAGAC	273	
Db	214 CGTGCACATCTCTGCTCAGCTGGGACAGTGGCCATACACCGAATTCCTCTTCCGACAGAC	273	
	477 GlnIleSerLeuValGlnTyrSerArgAspProHisMetGluPheSerLeuAsnArgTyr	496	
QY	AGCTCGGGGTGAGGCTGCCAGATCCGGATCCGGTCTTCTTGCCACGACGATGGTACACC	333	
Db	274 AGCTCGGGGTGAGGCTGCCAGATCCGGATCCGGTCTTCTTGCCACGACGATGGTACACC	333	
	497 AsnArgValLysAspIleIleGlnAlaIleAsnThrPheProTyrArgGlySerThr	516	
QY	CACACTGCTCTGGCGCTGCTGCTTACCAAGAACAGCTGTTTGTGAAGCATTCAGTGC	393	
Db	334 CACACTGCTCTGGCGCTGCTGCTTACCAAGAACAGCTGTTTGTGAAGCATTCAGTGC	393	
	517 AsnThrGlyLysAlaMetThrTyrValArgGluLysValPheValThrSerLysGlySer	536	
QY	CGGCGACGGGCTGCCAAAGTGTGTGTGGGAGACAGATGGGGCTCCAGCGACCTGTG	453	
Db	394 CGGCGACGGGCTGCCAAAGTGTGTGTGGGAGACAGATGGGGCTCCAGCGACCTGTG	453	
	537 ArgProAsnValProArgValMetIleLeuIleThrAspGlyLysSerSerAspAlaPhe	556	
QY	GGCCCCCATATGCAGAGCTCAAGAACCTGGGGCGTCACCGGTTCATTGTTCAGCACCGGC	513	
Db	454 GGCCCCCATATGCAGAGCTCAAGAACCTGGGGCGTCACCGGTTCATTGTTCAGCACCGGC	513	
	557 LysGluProAlaIleLysLeuArgAspAlaAspValGluIlePheAlaValGlyValLys	576	
QY	CGAGGCAACTTCCTGAGCTGTCAACCGCTGCTCAGCCCTTCGCGAGAACACCTGCAC	573	
Db	514 CGAGGCAACTTCCTGAGCTGTCAACCGCTGCTCAGCCCTTCGCGAGAACACCTGCAC	573	
	577 AspAlaValArgThrGluLeuGluAlaIleAlaSerProProAlaGluThrHisValTyr	596	
QY	TTTGTG---GAGCTGATGATCCTGCACATCATTTGTCCAAGACTGAGGGCTTCATTCTC	630	
Db	574 TTTGTG---GAGCTGATGATCCTGCACATCATTTGTCCAAGACTGAGGGCTTCATTCTC	630	
	597 ThrValGluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerValCys	616	
QY	GGAGTGGGGCGGACAGCTCCATGCC---	657	
Db	631 GGAGTGGGGCGGACAGCTCCATGCC---	657	
	617 LeuArgIleGluGlnGluLeuAlaAlaIleArgLysLysSerTyrValProAlaLysAsn	636	
QY	-----ACGAGATCATCGTCCACCGGGCTTCCGCTGAGCTGCAGCCACCCCTGCTGAC	708	
Db	658 -----ACGAGATCATCGTCCACCGGGCTTCCGCTGAGCTGCAGCCACCCCTGCTGAC	708	
	637 MetValPheSerAspValThrSerAspSerPheLysValSerTyrSerAlaIaGlySer	656	
QY	GCAGACTCGGCTACTATGTGCTGAGCTGTGTGCCACGCGCCACGCGGGGCTGCAGA	768	
Db	709 GCAGACTCGGCTACTATGTGCTGAGCTGTGTGCCACGCGCGCCACGCGGGGCTGCAGA	768	
	657 GluGluLysSerTyrLeuIleLysTyrLysVal-----AlaIleGlyLysAspGluPhe	674	
QY	CGCCACAGACGTCCGACGGAGAACCCACGACGTGATCTGGCGCGGCTGCACCCGACAG	828	
Db	769 CGCCACAGACGTCCGACGGAGAACCCACGACGTGATCTGGCGCGGCTGCACCCGACAG	828	
	675 IleValSerValProAlaSerSerThrSerSerValLeuThrAsnLeuLeuProGluThr	694	
QY	GACTACGACGTGACGCTAGTGCCTGATTCACAGCTGCGGCTCTGAGCGCCACAGATCTCG	888	
Db	829 GACTACGACGTGACGCTAGTGCCTGATTCACAGCTGCGGCTCTGAGCGCCACAGATCTCG	888	
	635 ThrTyrAlaValSerValIleAlaGlu-----Tyr	704	
QY	CGGGTGGCGCACCGGCGCCAGAGAGGCGGGCCAGAGCGCATTCACCGACCGCCG	948	
Db	889 CGGGTGGCGCACCGGCGCCAGAGAGGCGGGCCAGAGCGCATTCACCGACCGCCG	948	
	705 GluAspGlyAspGlyProProLeuAspGlyGluGluThrThrLeuGluValLysGlyAla	724	
QY	CGGCGACGCTCTCCG-----GTAGTTGGGCGCCA	978	
Db	949 CGGCGACGCTCTCCG-----GTAGTTGGGCGCCA	978	
	725 ProArgAsnLeuArgIleThrAspGluThrThrAspSerPheIleValGlyTyrThrPro	744	
QY	GGCGTGGGCTGACCGCGCGCGCTGCGGCTTACACGACGATTCGGGCGCTGCGGGGCGGG	1033	
Db	979 GGCGTGGGCTGACCGCGCGCGCTGCGGCTTACACGACGATTCGGGCGCTGCGGGGCGGG	1033	
	745 AlaProGlyAsn-----ValLeuArgTyrArgLeuValTyrArgProLeuThrGlyGly	762	
QY	GAGGCGACGCGGCTGAGAGTGGCCCGGGCCGCAACTGCACACGCTGCAGGGGCTGGCG	1091	

Db 763 GIUARGARGlnValThrValSerAlaAsnIuIuSerThrThrLeuArgAsnLeuIle 782
QY 1099 CCGGACCGCCCTACCTGCTGACCGCGACCGCCGCGCTTCCGCTCGGCGCGAGACGCG 1158
Db 783 ProAspThrArgTyrGluValSerValIleIleIuTyrGlnSerGlyProGlyAsnAla 802
QY 1159 CMTCCGCCAAGCGCTCGACGCCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1206
Db 803 LeuAsnGlyTyrAlaIuYsthr--AspGluValAlaGlyAsnProArg 817
RESULT 2
A54849
collagen alpha 1 (VII) chain precursor - human
N:Alternate names: procollagen alpha 1 (VII) chain
C:Species: Homo sapiens (man)
C:Date: 04-Nov-1994 #sequence, revision 04-Nov-1994 #text change 09-Jul-2004
C:Accession: A54849; PH0844; S16316; I56328; A30296; I84686
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.
J. Biol. Chem. 269, 20256-20262, 1994
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha (VII)
A:Reference number: A54849; MUID:94327588; PMID:8051117
A:Accession: A54849
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2944 <CHR>
A:Cross-References: UNIPROT:Q02388; UNIPARC:UPI000017A138; GB:I02870; NID:9387124; PIDN:
R.Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A:Title: Molecular cloning and characterization of type VII collagen cDNA.
A:Reference number: PH0844; MUID:92231902; PMID:1567409
A:Accession: PH0844
A:Molecule type: mRNA
A:Residues: 'EFR',340-475, 'RALSTASHSTLCWRATRHPCNRGSHWTTPACEPCNRPASHRAARAG',524-528,'C',
A:Cross-References: UNIPARC:UPI000017A139; DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:
A:Experimental source: Keratinocyte
A:Note: the authors translated the codon ACC for residues 394 and 397 as Tyr
R:Parente, M.G.; Chung, L.C.; Rymaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A:Reference number: S16316; MUID:91334380; PMID:1871109
A:Accession: S16316
A:Molecule type: mRNA
A:Residues: 815-892,'E',894-1439 <PAR>
A:Cross-References: UNIPARC:UPI000016K722; GB:M65158; GB:S49017; NID:9180914; PIDN:AAA96
A:Experimental source: Keratinocyte
R:Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Pristayanh, P.S.; Cook, M.E.; Wright, J.
J. Invest. Dermatol. 99, 691-696, 1992
A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot
A:Reference number: I56328; MUID:93107742; PMID:1469284
A:Accession: I56328
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'EFR',372-517,'DV',520-540,'W',542-1255 <RBS>
A:Cross-References: UNIPARC:UPI000016B343; GB:S51236; NID:9262308; PIDN:AA24637.1; PID:
R:Selzler, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, R.E.
J. Biol. Chem. 264, 3822-3826, 1989
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase
A:Reference number: A30296; MUID:89139437; PMID:2537292
A:Accession: A30296
A:Molecule type: Protein.
A:Residues: 'A',1240-1246,'G',1248-1250,'XE',1253-1255,'Q',1257,'E',2032,'C',2034-2041,'
A:Cross-References: UNIPARC:UPI000017A133; UNIPARC:UPI000017A13B; UNIPARC:UPI000017A13C;
A:Note: two reported peptides cannot be reliably located
R:Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A:Reference number: I48103; MUID:93271985; PMID:8499916
A:Accession: I48103
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: RNA
A:Residues: 2395-2871,'S',2873-2944 <RE2>
A:Cross-References: UNIPARC:UPI000016K724; GB:I06862; NID:9388713; PIDN:AAA89196.1; PID:

R.Christiano, A.M.; Rynaenen, M.; Uitto, J.
 Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
 A>Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser sub
 A:Reference number: A55255; MUID:94224777; PMID:8170945
 A:Contents: annotation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C:Genetic8:
 A:Gene: GDB:COL7A1; EBR1; EBD1; EB
 A:Cross-references: GDB:128750; OMIM:120120
 A:Map position: 3p21.3-3p21.3
 A:Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis
 A:Note: there are 118 introns
 C:Complex: type VII collagen is probably a homotrimer
 C:Function:
 A:Description: structural component of extracellular polymer associated with anchoring
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol
 F:1-16/Domin: signal sequence #status predicted <SIG>
 F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
 F:17-1253/Domin: amino-terminal nonhelical #status predicted <NC1>
 F:36-201/Domin: von Willebrand factor type A repeat homology <VWA1>
 F:231-318/Domin: fibronectin type III repeat homology <FN1>
 F:327-413/Domin: fibronectin type III repeat homology <FN2>
 F:414-502/Domin: fibronectin type III repeat homology <FN3>
 F:508-593/Domin: fibronectin type III repeat homology <FN4>
 F:598-683/Domin: fibronectin type III repeat homology <FN5>
 F:686-771/Domin: fibronectin type III repeat homology <FN6>
 F:776-862/Domin: fibronectin type III repeat homology <FN7>
 F:864-952/Domin: fibronectin type III repeat homology <FN8>
 F:954-1045/Domin: fibronectin type III repeat homology <FN9>
 F:1052-1219/Domin: von Willebrand factor type A repeat homology <VWA2>
 F:1170-1172/Region: cell attachment (R-G-D) motif
 F:1189-1253/Region: cysteine/proline-rich
 F:1254-2783/Region: interrupted helical
 F:1334-1336/Region: cell attachment (R-G-D) motif
 F:2008-2010/Region: cell attachment (R-G-D) motif
 F:2583-2555/Region: cell attachment (R-G-D) motif
 F:2784-2944/Domin: carboxyl-terminal nonhelical #status predicted <NC2>
 F:2876-2929/Domin: animal Kunitz-type proteinase inhibitor homology <BRI>
 F:337-786,1109/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status expe
 F:2655,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:2655,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental
 F:2654,2802,2804/Dissulfide bonds: Interchain #status predicted

Alignment Scores:
 Pred. No.: 5,44e-15 Length: 2944
 Score: 387.00 Matches: 137
 Percent Similarity: 41.9% Conservative: 47
 Best Local Similarity: 31.2% Mismatches: 205
 Query Match: 16.3% Indels: 50
 DB: 2 Gaps: 12

US-10-699-035A-5 (1-1254) x A54849 (1-2944)
 QY 31 CTGAGCTTGGCGGCTGGCGCGGAGCGCGCGGCTGCACGACATCAGCC 90
 Db 1 MetthrleuargleuValAlaAlaIleuGysAla-----GlyIleleuAlaGluAla 18
 QY 91 CCCCCA-----GGGAC 102
 Db 19 ProargValArgAlaGlnHisArgGluArgValThrCysThrArgLeuTyrAlaAlaSer 38
 QY 103 CTGATGTTCTGCTGCAGACAGCTGCAGACAGCTCTCACTAGAGTTCTCCGGGTTTCG 162
 Db 39 IlleValheuleuLeuAlaPrgLysSerSerIleLylArgSerAanpheArgGluValArg 58
 QY 163 GAGTTTGCGGCGGCGCTGGTGCCTCACTGCGCCCTGGGACCGGGGCC-----CTGCGT 216
 Db 59 SerpheuGluGluGluValLeuProPheSerGlyAlaAlaSerAlaGlnGlyValArg 78
 QY 217 GCACGTCTGGTGCACGTGGGAGTCCGCGCATACACGAGTTCCCTTCGGCAGACACAGC 276

```
D 79 PheAlaThrValGlnTyrSerAspAspProAlaArgThrGluPheGlyLeuAspAlaLeuGly 98
Q 277 TCGGGGTGAGGCGTCCCGAGATCGCGTGGCTCTTCTGCCACGCAATGGTGACACCCAC 336
D 99 SerGlyGlyAspValIleArgAlaIleArgGluLeuSerTyrIleGlyIleAsnThrArg 118
Q 337 ACTGGCTTGGCGCTGGTCTATGCTCCAGAGAACAGCTGTTCGTGAGACATCAGAGTCCCGG 396
D 119 ThrGlyAlaAlaIleLeuHisValAlaAspHisValPheLeuProGluLeu---AlaArg 137
Q 397 CCAGGGGTGCTCCAAAGTGTGGTGGTGGAGATGGCGGCTCCAGCAGCCCTGGGCG 456
D 138 ProGlyValProIlyValCysIleLeuIleThrAspGlyIlySerGlnAspLeuValAsp 157
Q 457 CCCCCCAGCAGAGCTCAAGAGCCTGGGCGTCAACCGTGTTCATTTGACACCGGCGCA 516
D 158 ThrAlaAlaGlnArgLeuIlyGlyGlnGlyValIlyLeuPheAlaValGlyIleIysAsn 177
Q 517 GGCMACTTCTGGAGCTGTTCAGCCGCTCAGCCCTGCGGAGAGACACCTGCACCTT 576
D 178 AlaAspProGluGluLeuIlyAspValAlaAspGlnProThrSerAspPhePhePhe 197
Q 577 GTGGAAGTGTGACCTGCACATCATT-----GNC 606
D 198 -----ValAsnAspPheSerIleLeuArgThrLeuLeuProLeuValSerArgVal 215
Q 607 CAAGAGCTGAGGGGCTCATCTCGGAGTCG-----CCG 642
D 216 CysThrThrAlaGlyIlyValProValThrArgProProAspAspSerThrSerAlaPro 235
Q 643 CAGCAGCTTCATGCGCAGCAGATCAGTTCAGAGGCTTCGCTCGCTCGGCCACCCCTG 702
D 236 ArgAspLeuValLeuSerIlyProSerSerGlnSerLeuArgValGlnThrAlaAla 255
Q 703 CTGACCGCAGACCTCGGCTACTATGTG-----CTGAGCTGGTGCCAGCGCCGCG 756
D 256 SerGlyProValThrGlyTyrIlyValGlnTyrThrProLeuThrGlyLeuGlyIlePro 275
Q 757 GGGGCTGCAAGACGCCAGAG-----CTGCCAGGAGAACGCCACGACATCTGGGCG 810
D 276 LeuProSerGluArgGlnGluValAsnValProAlaGlyIlyThrSerValArgLeuArg 295
Q 811 GGCCTGACCCGCGACGCACTACAGCGTGGCGCTGATGCTCAGTCCAACTGCGCCCTC 870
D 296 GlyLeuArgProLeuThrGluTyrGlnValThrValIleAlaLeuTyrAlaAsnSerIle 315
Q 871 CTGAGGCCCCAGATCCTGCGGGTGGCGCAGCGCGCAGAGGCGCGGCGAGCGCATC 930
D 316 GlyIlyAlaValSerGlyThrAlaArgThrThrAlaLeuGlu---GlyProGlu---Leu 333
Q 931 GTTCATCTCCACGCGCGCGCGCGCAGCTCCGCGTGAATGGAGCCCGACGCTGGAGCTCA 990
D 334 ThrIleGlnAsnThrThrAlaHisSerLeuLeuValAlaTyrArgSerValProGly--- 352
Q 991 GCCCGCGCGCTGGCTACACGTCGAGTTGCGGCGCGCTGCGGGGCGGGAGCGCGACGG 1050
D 353 -----AlaThrGlyTyrArgValThrTyrArgValLeuSerGlyGlyProThrGlnGln 370
Q 1051 GTGAGGTGCGCGCGCGCGCAACTGACACAGCGCTGAGCGCTGGCGCGGGACCGCG 1110
D 371 GlnGluLeuGlyProGlyGlnGlySerValLeuLeuArgAspLeuGluProGlyIlyThrAsp 390
Q 1111 TACCTGGTGAACCGTGAACCGCGCTTCCGCTCGGGCGCGGAGACCGCGTCTCCGCAAG 1170
D 391 TyrGluValThrValSerThrLeuPheGlyArgSerValGlyProAlaThrSerLeuMet 410
Q 1171 GCCTGACGCGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 1227
D 411 AlaArgThrAspAlaSerValGluGlnThrLeuArgProValIleLeuGlyProThr 429
RESULT 3
S31212
collagen alpha 1(XIV) chain precursor, short form - chicken
```

```
C/Species: Gallus gallus (chicken)
C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 15-Sep-2003
C/Accession: S31212
R/Waechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A/Title: Complete primary structure of chicken collagen XIV.
A/Reference number: S31211; MUID:93185668; PMID:8444186
A/Accession: S31212
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: mRNA
A/Residues: 1-1857 <MSE>
A/Cross-references: UNIPARC:UPI000017123C; EMBL:X70792; NID:9288874; PIDN:CA50063.1; PI
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C/Genetics:
A/Gene: Col14A1
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F1-28/Domin: signal sequence #status predicted <SIG>
F129-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>
F129-110/Domin: fibronectin type III repeat homology <FN3A>
F156-320/Domin: von Willebrand factor type A repeat homology <VWA1>
F1352-433/Domin: fibronectin type III repeat homology <FN3B>
F1442-525/Domin: fibronectin type III repeat homology <FN3C>
F1534-614/Domin: fibronectin type III repeat homology <FN3D>
F1623-707/Domin: fibronectin type III repeat homology <FN3E>
F1741-823/Domin: fibronectin type III repeat homology <FN3F>
F1832-914/Domin: fibronectin type III repeat homology <FN3G>
F1922-1009/Domin: fibronectin type III repeat homology <FN3H>
F1040-1205/Domin: von Willebrand factor type A repeat homology <VWA2>

Alignment Scores:
Pred. No.: 1,1e-14 Length: 1857
Score: 382.00 Matches: 118
Percent Similarity: 43.8% Conservative: 61
Best Local Similarity: 28.9% Mismatches: 158
Query Match: 16.1% Indels: 72
DB: 2 Gaps: 13

US-10-699-035A-5 (1-1254) x S31212 (1-1857)
Q 79 CCAGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 138
D 154 ProAlaIleAla-----AspIleValIleLeuValAspGlySerThrSerIleGly 170
Q 139 CACTACAGTGTCTCCGGGTTGCGGAGTTTGTGGGCGACGTGGCTCCACTGCGCCCTG 198
D 171 ArgPheAsnPheArgLeuValArgLeuPheLeuGluAsnLeuValSerAlaPheAsnVal 190
Q 199 GGCAGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258
D 191 GlySerGluTyrThrArgValGlyLeuAlaGlnTyrSerGlyAspProArgIleGluTyr 210
Q 259 CCCTTCCGCGCAGCAGACGCTGGGCTGGGCTGCCAGATGCGGTCGCTTCTCCCGAG 318
D 211 HisLeuAsnAlaTyrGlyThrIlyAspAlaValAlaLeuAspAlaValArgAsnLeuProTyr 230
Q 319 CGCATGGGTGACACCCACATGCGCTGGCGCTGTATGCCAGAAACAGCTGTTGCT 378
D 221 IysGlyGlyAsnThrLeuThrGlyLeuAlaLeuThrTyrIleLeuGluAsnSerPheIys 250
Q 379 GAAAGATACAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438
D 251 ProGluAlaGlyAlaArgProGlyValSerIlyIleGlyIleLeuIleThrAspIlyIys 270
Q 439 TCCAGACGCGCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 498
D 271 SerGlnAspAspValIleProProAlaIysAsnLeuAlaGAspAlaGlyIleIleLeuPhe 290
Q 499 ATTGTACAGACCGCGCGCGCACTTCTGAGACTGTACAGCGCTGACCGCTGCGC 558
D 291 AlaIleGlyValIlyAsnAlaAspIleAsnIleLeuIlyGluIleAlaSerGluProAsp 310
Q 559 GAAAGCACCTTCACCTTGTG---GACGTGATGACCTGACATCATTTGCGAA----- 609
```


QY 721 TACTATGCTGAGAGTGGTGGCCAGCCGCGGGGCTGCAAGACCCAGCAGCTG 780
Db |||||
387 TTTT-----Prothra3gylgylhprogluValValVal 400
QY 781 CCAGGAACGCGACGAGCTGATCTGGCGCGGACCCGACGAGCTACGACGTG 840
Db |||||
401 AepglYSerSerSerThrAlaValLeuYsAnleuMeSerLeuThrIglYrGlnle 420
QY 841 GCGCTAGT-----CCTGAGTCCAAC 861
Db |||||
421 AlaValAlaPheAlaIleTyrSerAnalaIaSerGluGlyLeuArgGlyThrGluThr 440
QY 862 CTGGCCCTCTCAGAGCCCAAGATCCTGGCGGGTCCGACCGCCGACGAGAGCCGCGCA 921
Db |||||
441 LeuAlaLeu-----PrometAlaSerAspLeu 449
QY 922 GAGCGCATGTGATCTCCAGCGCGCGCGCGAGCTCGCGTGTGGTGGCCCGACG 981
Db |||||
450 LysLeuTyrAspValSerHisSer-----SerMetArgAlaLysTTPasn----- 464
QY 982 CTGGGCTCAGCCGCGCGCTCGGCTACCACTGACAGTTCGGCGCGCGCGG--- 1038
Db |||||
465 ---GlyValAlaGlyAlaThrGlyTyrMetIleLeuTyrAlaProleuThrGluGlyLeu 483
QY 1039 GAGCGCGACGGGTGAGGTGCCCGCGCGCGCGACCTGACCACTGCTGAG-----GGC 1092
Db |||||
484 AlaAlaAspGluLysleuIleLysIleGlyLysAlaSerThrGluLeuLysAspGly 503
QY 1093 CTGGCGCGCGGACCGCTACCTGAGCGGTGACGTCGACCGCGCTCGCGCGCGCGAG 1152
Db |||||
504 LeuLeuProAsnThrGluTyrThrValThrValTyrAlaMetPheGlyLysLys 523
QY 1153 AGCGCGCTGTCCGCGCAAGGCTGACG 1179
Db |||||
524 AspProleuThrGlyGlnGluThrThr 532

RESULT 5
collagen alpha 1(XIV) chain precursor, short form 2 - chicken
NAlternate names: undulin
CSpecies: Gallus gallus (chicken)
CDate: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
CAccession: A45974; S30085; S22916; S17035; S20833
R(Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gemmari, M.; Dublet, B.; Cancedda, R.; Lin
J. Biol. Chem. 268, 12177-12184, 1993
A>Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region
na.
A.Reference number: A45974; MUID:93280195; PMID:8505337
A.Accession: A45974
A.Status: preliminary
A.Molecule type: mRNA; protein
A.Residues: 1-1747 <GER>
A.Cross-references: UNIPROT:P32018; UNIPARC:UPI0000173C47
A.Experimental source: embryo skin
A>Note: sequence inconsistent with the nucleotide translation
R(Ape, S.S.
submitted to the EMBL Data Library, March 1992
A.Reference number: S30085
A.Accession: S30085
A.Molecule type: mRNA
A.Residues: 1472-1660 <APT>
A.Cross-references: UNIPARC:UPI0000006A2; EMBL:X65122; NID:g62871; PIDN:CAA46238.1; PID
R(Trub, J.; Trub, B.
Eur. J. Biochem. 207, 549-557, 1992
A>Title: Type XIV collagen is a variant of undulin.
A.Reference number: S22916; MUID:92339443; PMID:1339349
A.Accession: S22916
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>
A.Cross-references: UNIPARC:UPI0000173C48
R(Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsenmayer, T.F.; van der Rest, M.; Mayne,

Eur. J. Biochem. 201, 333-338, 1991
A>Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens
A.Reference number: S17035; MUID:92037585; PMID:1935930
A.Accession: S17035
A.Molecule type: mRNA
A.Residues: 1472-1659 <GOR1>
A.Cross-references: UNIPARC:UPI0000173C49
A.Accession: S20833
A.Molecule type: protein
A.Residues: 1551-1570; 1593-1599; 1639-1667 <GOR2>
A.Cross-references: UNIPARC:UPI0000173C49
A.Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F/40-204/Domain: von Willebrand factor type A repeat homology <VWA1>
F/236-317/Domain: fibronectin type III repeat homology <FN3A>
F/326-409/Domain: fibronectin type III repeat homology <FN3B>
F/507-591/Domain: fibronectin type III repeat homology <FN3C>
F/625-707/Domain: fibronectin type III repeat homology <FN3D>
F/716-798/Domain: fibronectin type III repeat homology <FN3E>
F/806-893/Domain: fibronectin type III repeat homology <FN3F>
F/924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>
F/1111-1352/Domain: non-collagenous NC4 #status predicted <NC4>
F/1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>
F/1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Alignment Scores:
Pred. No.: 4,71e-14 Length: 1747
Score: 371.00 Matches: 116
Percent Similarity: 43.54 Conservative: 62
Best Local Similarity: 28.48 Mismatches: 159
Query Match: 15.64 Indels: 72
DB: 2 Gaps: 13

US-10-699-035A-5 (1-1254) x A45974 (1-1747)

QY 79 CCAGATAGACCCCGGAGGAGGAGCTGATGTTCTCTGTCGACAGTCCAGCGCTCTCT 138
Db |||||
38 ProAlaIleAla-----AspIleValIleLeuValAspGlySerTTPSerIleGly 54
QY 139 CACTACGAGTTCTCCCGGTTTCGGAGGTTTGGGGGCGCTGTTGCTCCACTGCGCCCTG 198
Db |||||
55 ArgPheAnpPheArgLeuValArgLeuPheLeuGluAnleuValSerArgPheAnVal 74
QY 199 GGCACCGGGCGCCCTGTCGACAGTCTGTGTGACGTCGGCGACGTCCGCTACCGAGTTC 258
Db |||||
75 GlySerGluLysThrArgValGlyLeuAlaGlnTyrSerGlyAspProArgIleGluThr 94
QY 259 CCCTTGGCGACGACAGCTCGGAGGCTGCCAGATGGGTCGCGCTTTCGCCAG 318
Db |||||
95 HisLeuAnlaTyrGlyThrLysAspAlaValLeuAspAlaAspArgAsnLeuProTyr 114
QY 319 CGCATGGGTGACACCACTGAGCTGCGCTGCTGTGCTTGCACAGACAGCTTTGCT 378
Db |||||
115 LysGlyGlyAsnThrLeuThrGlyLeuAlaLeuThrTyrIleLeuGluAnCysPheLys 134
QY 379 GAAGCATGAGTCCCGCGACGAGGGTCCCAAGTGTGTGTGGTGACAGATGCGCGC 438
Db |||||
135 ProGluAlaGlyAlaArgProGlyValSerLysIleGlyIleLeuIleThrAspGlyLys 154
QY 439 TCCAGCGACCTGTGGGCGCCCATGCAAGCTCAAGACCTGGGCGCTCACCGTTC 498
Db |||||
155 SerGlnAspAspValIleProProAlaLysAsnLeuArgGluAlaGlyIleGluLeuPhe 174
QY 499 ATTGCACACCGCGCGGCGAGCACTTCTGGAGCTGTGACGCGCTCGCTCACCGCTCC 558
Db |||||
175 AlaIleGlyValLysAsnAlaAspIleAsnGluLeuLysGluIleAlaSerGluProAsp 194
QY 559 GAGAAGCACTGCACTTTGTG---GACGTGATGACCTGCAATCATTTGTCCA----- 609
Db |||||
195 SerThrIleValTyrAsnValAlaAspPheAsnPheMetAsnSerIleValIleGluLeu 214
QY 610 -----GAGCTGAGGGGCTCCATTCTC 630
Db |||||

Db 215 ThrArgThrValCysSerArgValGluGluGluGluLysGluLysGlyThrIleAla 234
 QY 631 GCGATG-----CGCGCCGACGACTTCATGCGACGAGATCACTCCAGCGCTTCGCG 684
 Db 235 AlaSerLeuGlyAlaProThrAspLeuValThrSerAspIleThrAlaArgGlyPheArg 254
 QY 685 CTGGCCCTGGCCACCCCTGCTGACCGGACGACTCGGCG----- 720
 Db 255 ValSerTrp-----ThrHisSerProGlyLysValGluLysTrpArgValVal 270
 QY 721 TACTATGTGTGAGCTGTGTGCGACGCGCGCGGCTGCAGACGACGCGAGCTG 780
 Db 271 TyrTrp-----ProThrArgGlyGlyGlnProGluGluValVal 284
 QY 781 CCAGGAGACCGACGAGCTGATCTGGCGCGCGCTGCACCGGACGACGACTACGAGCTG 840
 Db 285 AspGlySerSerSerThrAlaValLeuLysAsnLeuMetSerLeuThrGluTrpGlnIle 304
 QY 841 GCGCTAGTG----- 861
 Db 305 AlaValPheAlaIleTyrSerAsnAlaAlaSerGluGlyLeuArgGlyThrGluThr 324
 QY 862 GTGCGCTCTCTGAGCGCCGACGATCTGCGGCTGCGACGCGCGACGAGAGCGCGGCCA 921
 Db 325 LeuAlaLeu-----ProMetAlaSerAspLeu 333
 QY 922 GAGCGCATCTCATCTCCACGCGCGCGCGCGACGCTCCGCGTGAAGTGGCGCCGACG 981
 Db 334 LysLeuTyrAspAlaSerHisSer-----SerMetArgAlaLysTrpAsn----- 348
 QY 982 CTGGGCTCAGCGCGCGCGCTGCGCTACACGTCGACAGTTCGGCGCGCTGCGCGCGCG 1038
 Db 349 ---GlyValAlaGlyAlaThrGlyTyrMetIleLeuTyrAlaProLeuThrGluGlyLeu 367
 QY 1039 GAGCGCAGCGGCGTGAAGGCGCGCGCGCGCGACGTCGACGACGCTGCGAG-----GGC 1092
 Db 368 AlaAlaAspGluLysGluIleLysIleGlyGlnAlaSerThrGluLeuGluAspGly 387
 QY 1093 CTGGCGCGCGGACCGCGCTACCTGCTGACCGGACGCGCGCTTCGCTGCGCGCGCGAG 1152
 Db 388 LeuLeuProAsnThrGluTrpThrValThrValTyrAlaMetPheGlyGluGluAlaSer 407
 QY 1153 AGCGCGCTGTCCGCAAGCGCTGCACG 1179
 Db 408 AspProLeuThrGlyGluGluThrThr 416
 RESULT 6
 151027
 type XII collagen alpha-1 chain - eastern newt (fragment)
 C:Species: Notoptthalmus viridescens, Triturus viridescens (eastern newt)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
 C:Accession: 151027
 R:Wei, Y.; Yang, E.V.; Klatt, K.P.; Tassava, R.A.
 Dev. Biol. 168, 503-513, 1995
 A:Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII collag
 A:Reference number: 151027; PMID:95246925; PMID:7729585
 A:Accession: 151027
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-929 <HEI>
 A:Cross-references: UNIPROT:Q91145; UNIPARC:UPI0000126D2E; EMBL:U19494; NID:9632647; PII
 F:155-236/Domain: fibronectin type III repeat homology <3FR>
 F:631-795/Domain: von Willebrand factor type A repeat homology <VMA3>

Alignment Scores:

Pred. No.: 2.16e-08 Length: 929
 Score: 272.50 Matches: 89
 Percent Similarity: 38.2% Conservative: 49
 Best Local Similarity: 24.7% Mismatches: 130
 Query Match: 11.4% Indels: 93
 DB: 2 Gaps: 9

US-10-699-035a-5 (1-1254) x 151027 (1-929)

QY 64 GCGACGCGGCTCA---CCAGCATGACC-----CCCGAGGGGAC 102
 Db 614 AlaProArgSerProIleProSerSerGlyLeuAspCysThrThrLysAlaGlnAlaAsp 633
 QY 103 CTGATGTTCTGCTGACAGACTGACGACGCTCTCAGTCTACAGATTCCTCCGGTTCCG 162
 Db 634 IleValLeuLeuValAspGlySerTrpSerIleGlyArgProAsnPheLysIleValArg 653
 QY 163 GAGTTGTGGGAGCGCTGCTGCACAGCCCTCGGACCGCGGCGCGCGCTGCGGACGT 222
 Db 654 AsnPheIleSerArgValValGluValPheAspIleGlySerAspArgValGlnIleAla 673
 QY 223 CTGTCACGTTGGCGACGTCGCGCATACACGAGTTCCTTCGCGACGACGCTGGGT 282
 Db 674 ValSerGlnTyrSerGlyAspProArgThrGluTrpGlnLeuAsnThrHisLysThrLys 693
 QY 283 GAGCGTCCGACAGATGCGGCTGCTGCTGCGCCACGCGACGTCGTCACACCCACGCGC 342
 Db 694 LysSerLeuMetAspAlaValAlaAsnLeuProTyrLysGlyLysAsnThrAsnThrGly 713
 QY 343 CTGGCGCTGCTCATGCTCAAGAAAGACGTTGCTGAAGCATCAGGTGCGCGCGACG 402
 Db 714 SerAlaLeuLysPheIleLeuGluGluAsnAsnPheArgProGlyValGlyMetArgGluLys 733
 QY 403 GTGCCAAAGTCTGTGCTGCTGCTGACAGATGCGGCTCCAGCGACCGCTGGGCGCCCGC 462
 Db 734 AlaArgLysIleAlaIleLeuLeuThrAspGlyLysSerGlnAspIleValAlaPro 753
 QY 463 ATGCAGAGACTCAAGACCTGGCGCTCACCCTGTTTCAATGTCAGCACCGGCGCGAC 522
 Db 754 SerLysArgTyrAlaAspGluGlyIleGluTyrAlaValGlyLysAsnAlaAsp 773
 QY 523 TTCCTGAGCTGTGACCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
 Db 774 GluAsnGluLeuLysGluIleAlaSerAspProAspGlu-----LeuTyrMetTyrAsn 791
 QY 583 GTGATGATCGTCGACATATGTCGAAGAGCTGAGGCGGCTCCATTCGCGATGCGCGCG 642
 Db 792 ValAlaAspPheSerLeuLeu----- 798
 QY 643 CAGCAGCTCCATGCCACGAGATCACGTCACGCGGCTTCGCGCTGCGCGCACCCCTG 702
 Db 799 -----ThrAsnIleValAsnAsp----- 804
 QY 703 CTGACCGCAGACTCGGGCTACTATGTCGTGAGCTGTGCGCGCGCGCGCGCGCGCT 762
 Db 805 -----LeuThrGluAsnValCysAsnSerValLysGly--- 815
 QY 763 GCMAAGCGCAGAGCTGCGCAGGAAAGCGCAGGACTGATGCTGCGCGCTGCAACCG 822
 Db 816 -----ProGly-----GlyLeuAsnPro 821
 QY 823 GACACGACTGACGACTGGCGCTAGTCTGAGTCCAGCGTCCGCTCGAGGCGCCAG 882
 Db 821 ----- 821
 QY 883 ATCTGCGGGTGCGACGCGGCGCAGAGCGCGCGCGCGCGCGCATCTCCAC 942
 Db 822 -----ProSerAsnLeuValThrSerGlu 929
 QY 943 GCGCGCGCGCAGACTCCGCGTGAATGGGCGCCCGCGCTGAGCGCGCGCGCTC 1002
 Db 830 ProThrProArgSerPheArgValThrTrpValProPro-----SerGlnSerValGlu 847
 QY 1003 GCGTACCATGTCGATGTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
 Db 848 ArgPheLysValGluTyrTyrProValAlaGlyArgProGlnGluValTyrValArg 867
 QY 1063 GCGGCGCGCAGTCGACGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
 Db 868 GlyThrGlnThrThrThrValLeuValGlyLeuLysProGluThrGluTyrTyrValAsn 887

```

QY      1123 GTG 1125
DB      888 Val 888

RESULT 7
undulin 1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: A40970
R:Just, M.; Herbst, H.; Hummel, M.; Duerkop, H.; Tripiier, D.; Stein, H.; Schnuppan, D.
J. Biol. Chem. 266, 17326-17332, 1991
A:Title: Undulin is a novel member of the fibronectin-tenascin family of extracellular m
A:Reference number: A40970; MUID:9137351; PMID:1716629
A:Accession: A40970
A:Molecule type: mRNA
A:Residues: 1-843 <US>
A:Cross-references: UNIPROT:Q05707; UNIPARC:UPI00000716A9; GB:M64108; NID:9340081; PIDD:
C:Keywords: glycoprotein
F:165-246/Domain: fibronectin type III repeat homology <FN3A>
F:255-338/Domain: fibronectin type III repeat homology <FN3B>
F:347-427/Domain: fibronectin type III repeat homology <FN3C>
F:436-520/Domain: fibronectin type III repeat homology <FN3D>
F:547-632/Domain: fibronectin type III repeat homology <FN3E>
F:641-723/Domain: fibronectin type III repeat homology <FN3F>
F:731-818/Domain: fibronectin type III repeat homology <FN3G>

Alignment Scores:
Pred. No.:      3, 94e-08      Length:      843
Score:          268.00        Matches:      97
Percent Similarity: 40.7%      Conservative: 53
Best Local Similarity: 26.3%      Mismatches: 149
Query Match:    11.3%         Indels:      70
DB:             2             Gaps:         13

US-10-699-035A-5 (1-1254) x A40970 (1-843)

QY      196 CTGGGACCGGGCCCTGCGTGCAGTGTGTCAGTGGGACGTCGACCGACCGAG 255
DB      3 ValGlySerGluThrArgIleGlyLeuAlaGlnTyrSerGlyAspProargIleGlu 22

QY      256 TTCCTCCCTGGCCAGACAGCTGGGTGAGGCTGCCAGAGTCCGCTGCTTGTGCC 315
DB      23 TRPHSLeuAsnAlaPheSerThrLysAspGluValIleGlnAlaValArgAsnLeuPro 42

QY      316 CAGCGCATGGGTGACACCCACACTGCGTGGCGCTGTCATGCCAAGAACAGCTGTT 375
DB      43 TyrLysGlyGlyAsnThrLeuThrGlyLeuAlaLeuAsnTyrIlePheGluAsnSerPhe 62

QY      376 GCTGAAGCATCAGTGCCTCCGCGGAGGGGTGCCAAAGTGTGCTGTGTCGACAGTGC 435
DB      63 LysProGluAlaGlySerArgThrGlyValSerLysIleGlyLeuIleThrAspGly 82

QY      436 GGCTCCAGGACCTGTGGGCCCCCCTCATTCAGAGACTCAAGACTCGGGCTCACCGTG 495
DB      83 LysSerGlnAspIleIleProSerArgAsnLeuArgGlnSerGlyValGluLeu 102

QY      496 TTCATGTGTGACGACCGCGGACCACTTCTGAGCTGTGCACCGCTCGTCAGCCCT 555
DB      103 PheAlaIleIleGlyValAsnAlaAspValAsnGluLeuGlnIleAlaSerGluPro 122

QY      556 GCCGAGACGACCTGTGCTTG---GACGTGATGACCTGCACATCTGTCAA--- 609
DB      123 AspSerThrIleValTyrAsnValAlaGluPheAspLeuMetHisThrValValGluSer 142

QY      610 -----GAGCTGAGGGCTCCATT 627
DB      143 LeuThrArgThrLeuCySerArgValGluGluGlnAspArgGlnIleLysAlaSerAla 162

QY      628 CTCGGCATGCGG-----CCGACAGAGCTTCATGACGACGATCCATCGTCAGCGGCTTC 681
DB      163 HisAlaIleThrGlyProProThrGlnLeuIleThrSerIleValThrAlaArgSerPhe 182

```

```

QY      682 CGCTGGCTGGCCACCCCTGCTACCGCAGACTCGGC----- 720
DB      183 MetValAsnTrp-----ThrHisAlaProGlnAsnValGlnLysTyrArgVal 198

QY      721 ---TACTATGTGCTGAGACTGTGTGTCACCGCCGCGGGGGGTGCAAGACGACGAG 777
DB      199 ValTyrTyr-----ProThrArgGlyLysLysProAspGluValVal 212

QY      778 CTGCCAGGAAACGCCACGACTGTGATCTGGGCGGGCTCGACCCGACACGACGACTACAC 837
DB      213 ValAspGlyThrValSerSerThrValLeuLysAsnLeuMetSerLeuThrGluTrgln 232

QY      838 GTGGGCTTATGTGCTGATTCAGTCCAGTGGCCCTCTCGAGGCCCCAGATCTCGGGTGGC 897
DB      233 IleAlaValPhe-----AlaIleTyrAlaHis 241

QY      898 ACGGCGCAGAGAG---GCCGCGCAGAGCGC----- 927
DB      242 ThrAlaSerGlnGlyLeuArgGlyThrGlnThrLeuAlaLeuProMetAlaSerAsp 261

QY      928 ATCGTCATCTCCACGCGCCGCGCAGCCTCCGCTGAGTTGGCCCGACCGCTGGGC 987
DB      262 LeuLeuLeuTyrAspValThrGluAsnSerMetArgValLysTrpAspAlaValProGly 281

QY      988 TCAGCCGCGCGCTGCTGCTACCACTGCACTTCGAGCGCGCTGCGGCGCGGAGCG--- 1044
DB      282 -----AlaSerGlyTyrLeuIleLeuTyrAlaProLeuThrGlnGlyLeuAlaGly 298

QY      1045 CAGCGGATGAGAGTGCCTCCGCGCGCGCAACTGACAC---ACGCTGACGAGCGCTGGCG 1098
DB      299 ArgGluLysGlnMetLysIleGlyGlnThrHisThrAspIleGlnLeuSerGlyLeuLeu 318

QY      1099 CCGGACACCGCTTACCTGATGTCACCGTGAACCGCGCTTCCGC----- 1140
DB      319 ProAsnThrGluTyrThrValThrValTyrAlaMetPheGlyGluAlaLaserAspPro 338

QY      1141 ---TCGGCGCGCAGAGCGCGCTGCC 1164
DB      339 ValThrGlyGlnGluThrThrLeuAla 347

RESULT 8
Q08B3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 09-Jul-2004
C:Accession: A03742
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825
A:Accession: A03742
A:Molecule type: DNA
A:Residues: 1-660 <BN>
A:Cross-references: UNIPROT:P03181; UNIPARC:UPI000000CD00
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667; PMID:6087149
A:Contents: annotation; protein coding region
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
C:Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
Pred. No.:      7, 8e-08      Length:      660
Score:          263.00        Matches:      166
Percent Similarity: 34.4%      Conservative: 28
Best Local Similarity: 29.4%      Mismatches: 197
Query Match:    11.1%         Indels:      173
DB:             1             Gaps:         32

US-10-699-035A-5 (1-1254) x Q08B3 (1-660)

QY      15 GCGCTCGGCTGGCCCTGAGCTTGCGGCTGGCGCGGAGCGCGGAGCGCGG 74

```


Score: 248.50 Matches: 66
 Percent Similarity: 49.4% Conservative: 82
 Best Local Similarity: 37.1% Mismatches: 23
 Query Match: 10.4% Indels: 7
 DB: 2 Gaps: 2

US-10-699-035A-5 (1-1254) x A33809 (1-493)

```

QY 73 GGTCCACACGATTCAGCCCCCGAGGG-----GACCTGATGTTCTCG 114
DB 257 GtlyvstthCysSerAlaCysSerGtlyGlySerGlySerAlaLeuAspLeuValPheLeu 276
QY 115 CTGAACAGCTCAGCCAGCGCTCTCTCACTACAGATTCTCCCGGATTCCGAGATTGTGGG 174
DB 277 lIeaRgIySerIySerValArgrProGluAsnPhnGluLeuValIySerPheIleaRn 296
QY 175 CAGCGTGGCTCCACTGCCCTGGGACCGGGGCCCTGCGCCAGCTGTGTCAGCTG 234
DB 297 GlnIleValGluSerLeuGluValSerGluIySgIlnAlaGlnValGlyLeuValGlnTyr 316
QY 235 GGCAGTCGGCCATACACGAGTTCCCTTCGGCCAGACAGCTCGGATGAGCTGCCAG 294
DB 317 SerSerSerValArGlnInIuPheProLeuGlnPheIyAsnIySAspIleIyS 336
QY 295 GATCGCGTCCGCTTCTCCAGCGCATGGGTGACACCCACACTGCGCTGGCGCTGTC 354
DB 337 AlaAlaValIyIySerMetAlaTyrMetGluIySgIyThrMetThrGlyGlnAlaLeuIyS 356
QY 355 TTTGCCAAGAACAGCTGTTTCTGTAAGATCAGTCCGGCCAGGGGCGGCGCAAGT 414
DB 357 TyrLeuValAspSerSerPheSerIleAlaAsnGlyAlaArgrProGlyProIyVal 376
QY 415 CTGTGGTGGGTGACAGATGCGGCTCCAGCGACCCCTGTGSGCCCCCATGACGAGCTC 474
DB 377 GtlyIleValPheThrAspGlyArGserGlnAspTyrIleThrAspAlaIySAspIyS 396
QY 475 AAGACCTGGGCGTCAACCGTGTTCATTGTCAGACACCGCCGAGGCACTCTTCGAGCTG 534
DB 397 IyAspIyLeuGlyPheArGmetPheAlaValGlyValGlyAsnAlaValGluAspGluLeu 416
QY 535 TCAGCGCGTCCGCTCAGCGCTGCCGAGAACCTTCGACTTTGTGACGTGAT 588
DB 417 ArGtGluIleAlaSerGluProValAlaGluHis--TyrPheTyrThraIAsp 433

```

RESULT 10

J00405
 hypothetical 119.5K protein (uvra region) - Micrococcus luteus
 N:Alternate names: ORF 1 protein
 C:Species: Micrococcus luteus, Micrococcus lysodeikticus
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
 C:Accession: J00405
 R:Shiota, S.; Nakayama, H.
 Mol. Genet. 217, 332-340, 1989
 A:Title: Micrococcus luteus homolog of the Escherichia coli uvra gene: identification of
 A:Reference number: S04781; MUID:89364717; PMID:2549377
 A:Accession: J00405
 A:Molecule type: DNA
 A:Residues: 1-1106 <SH1>
 A:Cross-references: UNIPARC:UPI0000177394; EMBL:X15867
 A>Note: this reading frame extends between two stop codons and does not begin with a sta
 A>Note: the gene encoding this protein overlaps uvra gene
 C:Superfamily: collagen alpha 1(1) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:

Pred. No.: 6.53e-07 Length: 1106
 Score: 246.50 Matches: 157
 Percent Similarity: 31.1% Conservative: 19
 Best Local Similarity: 27.7% Mismatches: 183
 Query Match: 10.4% Indels: 207
 DB: 2 Gaps: 29

US-10-699-035A-5 (1-1254) x J00405 (1-1106)

```

QY 24 CTTGGCCTTGAGCTTGGCGGCTGGCGCTGGCGCGGAGCGGCGC-----GAGCG 71
DB 569 ProGlyPro-----AlaGlyValProIaGlyArgrProArgrValProGlnPro 585
QY 72 CGGTCACACAGATCAGCGCC-----CGAGGGGACCTGATGTTCTGTGTCAGACGCTC 125
DB 566 ArGAlaAlaGlyArGnIePProLeuArgrArgrGlyProAlaIleArPProGlyIleThr 605
QY 126 AGCCAG--CGTCTCTCACTAGAGTTCTC--CGGGTTCCGGAGTTTGGGGCAGCT 179
DB 606 AsrArGleArGAlaGlyArGArgrProLeuArgrProArgrArGAlaValIleArgrProIa 625
QY 180 GGTGCTCACTACCTCCCGGACCGGGGCCCTGGCTGCTGTCAGCT----- 233
DB 626 ProAlaGlyGlnProProIleArgrArgrProArProAlaProAlaGlyProArgrGlnHis 645
QY 233 ----- 233
DB 646 ProIleArGrArGrAlaArGrArGlyIleIyAsrArGrArGlyGlyLeuAsrArGrGlyIleHis 665
QY 233 ----- 233
DB 666 ArgrProSerArgrGlyArGrValArgrArgrGlyArGrAlaLeuGlyLeuProGlyGlySer 685
QY 234 ---GGGAGTCGGCCATA-----CAGGAGTTCCCTTCGGCCAGCACAGCTCGGATGA 284
DB 686 GlnGlyIleIleIleValAlaArGnIeArGrArgrIleuProIeArgrPro----- 700
QY 285 GCGTCGCCA-----GATGCGGTGGCGCTTCTGTCGCCAGCGATGGGTGACAC-- 332
DB 701 -----ProIeIleArGrGlyAlaGlyAlaIleArGrArgrProGlyGlyIleValArGrAla 718
QY 333 -----CSACATGCGCTGCG-----CGTGTCTATGCSAAGAACAGCTGTT 374
DB 719 AsrGlyProArGrArgrProGlyGlnGlnIleProGlyGlyArGrLeuGlyProGly----- 735
QY 375 TGCTAAGCATAGGTGCGCGGCTCGAGGGGTGCGCCAAAGTCTGTGTGGGTGACAGATGG 434
DB 736 -----ProIaArGrGlyProIleHis-----GlyArGrAspGlyArGrVal 747
QY 435 CGGCTCCAG-----CGACCTGT-----GAGCCCCC 461
DB 748 ArGrLeuArGrGlnValIleAlaArGrIleArGrArPProLeuGlnGlyProGlyGlnProIa 767
QY 444 -----CGACCTGT-----GAGCCCCC 461
DB 768 GlnArGrArGlnAlaArGrAlaArgrProIaArGrValArGrGlyIyThraArGrAlarProGly 787
QY 462 CATGCAAGAGCTCAAGACCTGGCGGTACCGGTGTTCTGTGCAGACCGCGCGAGGCA 521
DB 788 GlnGlyGlyProArGrGlyProGlyProIleArGrAlaIleAlaThrLeuGlnProArGrHis 807
QY 522 CTTCTGAGAGCTGTCAGCGCGCTGCCCTCAGCGCCCTCCAGAAAGCACTGCACTTTGTGGA 581
DB 808 LeuIleSgIyArGrValArGrArgrArgrProGlnAlaIleAspArGrGlyLeuGly 827
QY 582 CGTGATGACCTGCA-----CATCATGTTCCAAAGCT---GAGGGGCTC 623
DB 828 ProGlyLeuProIaGlyProValLeuIleuGlnIleSglnGlyArGrAlaIleuArGrGlyVal 847
QY 624 CATTCCTCGGATCGCGC----- 641
DB 848 ArGtIyArGrArGnIeAlaGluAsrArGrArGrIyLeuProAlaGlyArGrLeuArGrAlaVal 867
QY 641 ----- 641
DB 866 ArGrIyValProArGrGlyProValGlnProGlyAsrAlaArGrGlyIleIleuGlnIyGln 887
QY 642 -----GAGCAGCTCCATGCGCACGAGATCAC-----GTCCAGCGGCTT 680
DB 888 GlnIleArGrArGrGlyProArGrIleIleAlaArGrArGrIyGlyArGrGlyLeuIleuGlnIyArGrVal 907
QY 681 CCGCCTGGCGCTTGCCACCCCTGCTGAC-----CGCAGACTCGGGCTACTATGTGCT 731

```


Db 908 HisProHisLeuValProGlyHisAlaArgArgArgSerGlyLeuArgProSer 927
732 GGA-----GCTGTGCTCCAGCGCCAGCGGGGCTGC 764
Qy 928 GlyProAlaArgHisHisAlaLeuGlyArgArgGlyProAlaArgGlyAlaGlyGly--- 946
Qy 765 AAGACGCCAGACCTCGCAGGAAAGCCAGCAGCTGAGTGGCCGCTCGACCCGGA 824
Db 947 ArgAlaProGluAlaLeuGlnArgProHisHisIleuArgAlaGlyAlaAspHisGly 966
Qy 825 CACGAGCTAGAGCTGGCCGCTAGTGGCTGAGTCCACAGTGCCTCTGAGGCCACAT 884
Db 967 ValAlaLeuArgArg-----HisProGlnAlaProAla 977
Qy 885 CCGCGGCTCGCAGCGCGGAGAGCGCGGCGCAGAGCGATGCTCATCTCCACGC 944
Db 978 ArgAlaSerValProArgGlyGlnGlyGlnHisGlyAlaHisHisArg----- 993
Qy 945 CCGGCGCGCAGCTCGCGCT-----GAGTTGGGCGCCAGCGCTGGGCTCAGCCGCGC 998
Db 994 -----AlaGlnProArgArgAspGlnGlnArgGlyProArgAspArgProArgProGly 1011
Qy 999 GCTCGGCTACACAGTGCAGTTCCGGCCGCTCGCGGCGGAGAGCCGAGCGGTGAGGT 1058
Db 1012 GlyArgLeuArgArg-----ArgHisAspArgGlyHisGlyHisAlaGlyGlyArg 1029
Qy 1059 GCGCGGCGGCGCAGCTGACACAGCTGACAGCGCTGCGCGGCGCAGCGCTGCTGCT 1118
Db 1030 AlaArgArgArgGluProHis-----GlyPro-ValProArgGlyAlaAlaArgVal 1046
Qy 1119 GACCGTACCGCGCGCTTCGCTCGCGGCGCGAGAGCGCGCTGCTCGCAGAGCGCTGAC 1178
Db 1046 AlaIleTrpThr-ProProThrGlnAspGlnAspGlnGlyGlu---ProTrpArgArgArg 1065
Qy 1179 GCGCGAGCG 1238
Db 1065 rgrProCysAlaGlyAla-----ProProArgCysCysGlySerp 1078
Qy 1239 CAGCGGTGAGCGGT 1252
Db 1078 rAlaValProArg 1082

RESULT 11
S66522
cartilage matrix protein precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C/Accession: S66522
R/Ascoli, A.; Hauser, N.; Studer, D.; Paulsson, M.; Hirtl, L.; Bosze, Z.
Eur. J. Biochem. 236, 970-977, 1996
A/Title: Cloning, sequencing and expression analysis of mouse cartilage matrix protein C
A/Reference number: S66522; MUID:96270751; PMID:8665920
A/Accession: S66522
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-500 <ASZ>
A/Cross-references: UNIPROT:P51942; UNIPARC:UPI000002480; EMBL:U35035; NID:91163178; PI
C/Genetics:
A/Genes: CNP
C/Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
F/1-29/Domain: signal sequence #status predicted <SIG>
F/30-500/Product: cartilage matrix protein #status predicted <MAT>
F/43-210/Domain: von Willebrand factor type A repeat homology <VMA1>
F/231-266/Domain: EGF homology <EGF>
F/277-441/Domain: von Willebrand factor type A repeat homology <VMA2>

Alignment Scores:
Pred. No.: 1.05e-06 Length: 500
Score: 243.50 Matches: 60
Percent Similarity: 50.9% Conservative: 29
Best Local Similarity: 34.3% Mismatches: 75
Query Match: 10.2% Indels: 11

DB: 2 Gaps: 2
US-10-699-035A-5 (1-1254) x S66522 (1-500)
Qy 91 CCGCGAGG-----GACGTAGTTCCTGCTGACAGCTCA 126
Db 34 ProArgGlyHisIleuCysArgThrArgProThrAspLeuValPheValAlaAspSer 53
Qy 127 GCCAGCGTCTCATACAGATTCCTCCGAGTTCCGAGAGTTTGGAGGAGCTGAGGCT 186
Db 54 ArgSerValArgProValGlnPheGlnValValValValPheLeuSerGlnValIleGlu 73
Qy 187 CCACTGCCCTTGCGACCGCGGCGCTGCTGCCAGTCTGTCAGCTGCGAGCTCGGCCA 246
Db 74 SerLeuAspValGlyProAsnAlaThrArgValGlyLeuValAsnTyraIaSerThrVal 93
Qy 247 TACACGAGTTCCTTCGCGCAGCAGCTGCGGTGAGGCTGCCAGATGCGGCGGT 306
Db 94 LysProGlnPheProLeuArgAlaHisGlySerLysAlaSerLeuLeuGlnAlaValArg 113
Qy 307 GCTTTCGCCAGCGCATGGGTGACACCCACACTGAGCTGCGGTGCTTATGCCAAGAA 366
Db 114 ArgIleGlnProLeuSerThrArgIleMetThrGlyLeuAlaLeuGlnPheAlaIleThr 133
Qy 367 CAGCTTTCTGACAGCATCAGGTCGCG-----CGAGGCTGCCAAGTCTG 417
Db 134 LysAlaLeuSerAspAlaGlnGlyArgAlaArgSerProAspIleSerLysValVal 153
Qy 418 GTCGTGGTACAGATGCGCGCTCCAGACCGCTGTCGCGCGCGCGCGCGCGCGCGCG 477
Db 154 IleValIleThrAspIleArgProGlnAspSerValArgAspValSerGlnArgAlaArg 173
Qy 478 GACCTGCGTACCGCTGTTTCATTTTCACAGCAGCGCGCGAGGAGCACTTCTGAGCTGCA 537
Db 174 AlaSerGlyIleGluLeuPheAlaIleGlyLeuGlyArgValAspLysAlaThrLeuArg 193
Qy 538 GCGGCTGCTCAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
Db 194 GlnIleAlaSerGlnProGlnAspGlnHisValAspTyraValGlu 208

RESULT 12
Q08B3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C/Species: human herpesvirus 4, Epstein-Barr virus
C/Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 09-Jul-2004
C/Accession: A03742
R/Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr viru
A/Reference number: A93065; MUID:85035713; PMID:6092825
A/Accession: A03742
A/Molecule type: DNA
A/Residues: 1-660 <BAN>
A/Cross-references: UNIPROT:P03181; UNIPARC:UPI000000CDDC
R/Baker, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; I
Nature 310, 207-211, 1984
A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A/Reference number: A03794; MUID:84270667; PMID:6087149
A/Contents: annotation; protein coding region
C/Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-5
C/Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
Pred. No.: 1.02e-06 Length: 660
Score: 243.50 Matches: 150
Percent Similarity: 32.8% Conservative: 25
Best Local Similarity: 28.1% Mismatches: 183
Query Match: 10.2% Indels: 175
DB: 1 Gaps: 34

US-10-699-035A-5 (1-1254) x Q08B3 (1-660)
Qy 1237 CGGTCCCGGGGTGGGGCGCGGGGCGAGGCGTGGGC-----GCGGCGCGGCGC 1187

A:Accession: A35887
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: UNIPARC:UPI0000174566; GB:J05582; NID:G189598; PIDN:AAA60019.1; PID:R:Wreschner, D.H.; Haruvirt, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
A:Title: Human epithelial tumor antigen CDNA sequences. Differential splicing may generate
A:Reference number: S10571; MWID:90276413; PMID:2351132
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A:Cross-references: UNIPARC:UPI0000174569; EMBL:X52229; NID:G37053
A:Reference number: UNIPARC:UPI0000174569; EMBL:X52229; NID:G37053
A:Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A:Reference number: S40293
A:Accession: S40293
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2>
A:Cross-references: UNIPARC:UPI000016B0A6; EMBL:X52229; NID:G37053; PIDN:CAA36478.1; PID:R:Abel, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MWID:90088473; PMID:2597151
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A:Cross-references: UNIPARC:UPI0000174564; EMBL:M31823; NID:G181542; PIDN:AAA35757.1; PID:R:Maunzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglut
A:Reference number: JX0235; MWID:93123189; PMID:1478919
A:Accession: JX0235
A:Molecule type: mRNA
A:Residues: 998-1011,'RS',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
A:Cross-references: UNIPARC:UPI0000174568; UNIPARC:UPI000017456C; UNIPARC:UPI000017456D
A:Experimental source: gastric carcinoma cell
R:Zirhan-Licht, S.; Baruch, A.; Eliroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
A:Reference number: S51026; MWID:95080414; PMID:7988707
A:Contents: annotation
A:Note: Undetermined tyrosine residues in the carboxyl-terminal non-repetitive region at
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 C
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Genetics:
A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-1q23
A:Introns: 20/1: 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
E:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRA>
F:1-62/Region: mucin 1 amino-terminal non-repetitive
F:1-23/Domain: signal sequence #link PRA #status predicted <SIGA>
F:1-19,29-1344/Domain: signal sequence #link PRA #status predicted <SIGB>
F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PRB>
F:1-19,29-21,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
F:1130-1017/Region: 20-residue repeats (GSTPAHGVTASPTRPAP)
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:1245-1372/Domain: transmembrane #status predicted <TMS>
F:1046,1064,1118,1144,1222/Binding site: carboxylate (Asn) (covalent) #status predicted
F:1123/Binding site: phosphate (Tyr) (covalent) #status predicted

Alignment Scores:

Pred. No.:	1,51e-06	Length:	1344
Score:	240.00	Matches:	140
Percent Similarity:	28.3%	Conservative:	29
Best Local Similarity:	34.3%	Mismatches:	189
Query Match:	10.1%	Indels:	136

[illegible]

```

Db      890 AlaProAspThrArgProAlaProGlySerThr---AlaProProAlaHisGlyValThr 908
QY      307 CACGACCGCAT-----CCTGGCAGACCTCACCGAGCTGTGCTGGCGAAG 260
Db      909 SerAlaProAspThrArgProAlaProGlySerThrAlaProProAla-----His 925
QY      259 GGAACCTCGGTATGATGCGCAGCTGCCACGACTGCGACGACGCGCGCGG--- 203
Db      926 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 945
QY      202 -----TCCCGAGGGGACGTGAGCCACCACT--- 176
Db      946 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 965
QY      175 -----GCCCGACAAACT---CCGAAACCGGAGAACTGCTGTAGTAGAGACGC 131
Db      966 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 985
QY      130 TGGCTGAGC-----TCTCCAGCAGGAACATCAGCTCCCTCGGG 92
Db      986 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 1005
QY      91 GGGCTGATGCTGTG---GACCGCGCTCGCGCGCTCGCGCCGACGCGCGCA--- 38
Db      1006 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 1025
QY      37 -----AGCTCAGGCGCCGAGCGCGCTCCAGGAGACA 2
Db      1026 GlyValThrSerAlaProAspThrArgProAlaProGlySer 1039

```

RESULT 14

cartilage matrix protein precursor - human

```

C/Species: Homo sapiens (man)
C/Date: 12-Jul-1991 #sequence_revision 12-Jul-1991 #text_change 09-Jul-2004
C/Accession: A37979; B37979
R/Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; SH
J. Biol. Chem. 265, 19624-19631, 1990
A/Title: Structure and chromosomal location of the human gene encoding cartilage matrix
A/Reference number: A37979; MUID:91060568; PMID:2246248
A/Accession: A37979
A/Molecule type: DNA
A/Residues: 1-496 <JEN>
C/Accession: B37979
A/Cross-references: UNIPROT:P21941; UNIPARC:UPI000004F1ED; GB:J05667
A/Accession: B37979
A/Molecule type: mRNA
A/Residues: 157-290, 'L', 292-496 <JEN>
A/Cross-references: UNIPARC:UPI000016A6E8; GB:M55683; GB:J05666; GB:J05667; NID:G180651;
C/Genetics:
A/Gene: GDB:CRIM
A/Cross-references: GDB:127280; OMIM:115437
A/Map position: 1p35-1p35
A/Introns: 32/1, 147/3; 222/1; 264/1; 403/1; 454/1; 481/1
C/Complex: homotrimer
C/Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
C/Keywords: glycoprotein; homotrimer
F.1-22/Domain: signal sequence #status predicted <Sig>
F.23-496/Product: cartilage matrix protein #status predicted <MAT>
F.39-206/Domain: von Willebrand factor type A repeat homology <WMA1>
F.227-262/Domain: EGF homology <EGF>
F.1273-437/Domain: von Willebrand factor type A repeat homology <WMA2>
F.76,344/Binding site: carbohydrate (Asn) (covalent) #status predicted
F.221-238,234-247,249-262/Disulfide bonds: #status predicted

```

Alignment Scores:

```

Pred. No.: 3,91e-06 Length: 496
Score: 233.50 Matches: 62
Percent Similarity: 50.3% Conservative: 27
Best Local Similarity: 35.0% Mismatches: 87
Query Match: 9.8% Indels: 1
DB: 2 Gaps: 1

```

US-10-699-035a-5 (1-1254) x A37979 (1-496)

```

QY      100 GACCTGATGCTTCCTGCTGAGACGCTCAGCCAGCGCTCTCTCACTAGCACTCTCCGGGTT 159
Db      275 AspLeuValPheLeuIleAspGlySerIleSerValArgProGluAsnPhelGluLeuVal 294
QY      160 CGGAGCTTTGGGGACACTGTGTGCTCCACTGCCCTCTGGGACCGGGGCTGTGCTGCC 219
Db      295 LysIleSperIleSerGlnIleValAspThrLeuAspValSerAspLysLeuAlaGlnVal 314
QY      220 AGCTGTGTCACGCTGGGACGTCGGCCATVACCCGAGTCCCTTCCGCGACACGACTCG 279
Db      315 GlyLeuValGlnIleTyrSerSerSerValArgGlnGluPheProLeuGlyArgPheIleThr 334
QY      280 GGTGAGGCTGCCAGGATGCGGTGCGCTCTTTCGCCAGCCAGTGGTACACCACT 339
Db      335 LysLysAspIleLysAlaIleValArgAsnMetSerTyrMetGluLysGlyThrMetThr 354
QY      340 GGCCTGGCGCTGTGTTATGCCAGGACAGCACTGTTCTGTAAGCATCAGTCGCGCGCA 399
Db      355 GlyAlaAlaLeuLysTyrLeuIleAspAsnSerPheThrValSerSerGlyAlaArgPro 374
QY      400 GGGGTGCCCAAGTCTGTGTGGGTGACAGATGAGCGGCTCCAGCGACCTGTGGGCCCC 459
Db      375 GlyAlaGlnLysValGlyIleValPheThrAspGlyArgSerGlnAspTyrIleAsnAsp 394
QY      460 CCCATGACAGAGCTCAGACGCTGGGCGTCAACCGTTTCATTGTGACGACCGCGGAGGC 519
Db      395 AlaAlaLysLysValAlaLysAspLeuLysPheLysMetPheAlaValGlyValGlyAsnAla 414
QY      520 AACTTCCTGAGCTGTACGCGCTGCTCAGCCCT---GCCGAAACAGCACTGACATT 576
Db      415 ValGluAspGluLeuArgGluIleAlaSerGluProValAlaGlnIleTyrPheTyrThr 434
QY      577 GTGACGTGATGACCTGACATCATTCGCAAGAGCTGAGGGCTCCATT 627
Db      435 AlaAspPheLysThrIleAsnGlnIleGlyLysLeuGlnLysLeuIle 451

```

RESULT 15

A37797
collagen alpha 3(VI) chain precursor - chicken

```

C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Aug-2004
C/Accession: A37797; A34270; A32674
R/Doliana, R.; Bonaldo, P.; Colombatti, A.
J. Cell Biol. 111, 2197-2205, 1990
A/Title: Multiple forms of chicken alpha3(VI) collagen chain generated by alternative sp
A/Reference number: A37797; MUID:91035630; PMID:1977751
A/Accession: A37797
A/Molecule type: mRNA
A/Residues: 1-253;312-321;434-453;504-518;635-655;704-717;832-853 <DOL>
A/Cross-references: UNIPROT:P15989; UNIPARC:UPI0000173C31; UNIPARC:UPI0000173C32; UNIPAR
:M24282
R/Bonaldo, P.; Russo, V.; Buccicotti, F.; Doliana, R.; Colombatti, A.
Biochemistry 29, 1245-1254, 1990
A/Title: Structural and functional features of the alpha3 chain indicate a bridging role
A/Reference number: A34270; MUID:90212613; PMID:2325559
A/Accession: A34270
A/Molecule type: mRNA
A/Residues: 224-2871 <BON>
A/Cross-references: UNIPARC:UPI0000173C38; GB:M24282
A/Note: the authors translated the codon TTC for residue 1916 as Leu and TTC for residue
R/Bonaldo, P.; Colombatti, A.
J. Biol. Chem. 264, 20235-20239, 1989
A/Title: The carboxyl termini of the chicken alpha3 chain of collagen VI is a unique mo
A/Reference number: A32674; MUID:90062147; PMID:2584214
A/Accession: A32674
A/Molecule type: mRNA
A/Residues: 2151-2199;2792-3137 <BO2>
A/Cross-references: UNIPARC:UPI0000173C39; UNIPARC:UPI0000173C3A; GB:M24282
C/Genetics:
A/Introns: 30/1, 236/1, 437/1; 638/1; 838/1
C/Superfamily: collagen VI

```

C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; extracellular
F:125/Domain: signal sequence #status predicted <SIG>
F:26-3137/Product: collagen alpha 3(VI) chain #status predicted <MAT>
F:26-2042/Domain: non-collagenous #status predicted <NMC>
F:36-202/Domain: von Willebrand factor type A repeat homology <VM01>
F:229-404/Domain: von Willebrand factor type A repeat homology <VM02>
F:442-607/Domain: von Willebrand factor type A repeat homology <VM03>
F:642-807/Domain: von Willebrand factor type A repeat homology <VM04>
F:840-1004/Domain: von Willebrand factor type A repeat homology <VM05>
F:1033-1197/Domain: von Willebrand factor type A repeat homology <VM06>
F:1237-1400/Domain: von Willebrand factor type A repeat homology <VM08>
F:1439-1604/Domain: von Willebrand factor type A repeat homology <VM09>
F:1639-1804/Domain: von Willebrand factor type A repeat homology <VM10>
F:1838-2010/Domain: von Willebrand factor type A repeat homology <VM11>
F:2043-2378/Domain: collagenous #status predicted <COL>
F:2045-2047/Region: cell attachment (R-G-D) motif
F:2153-2155/Region: cell attachment (R-G-D) motif
F:2159-2161/Region: cell attachment (R-G-D) motif
F:2379-3137/Domain: non-collagenous #status predicted <CNC>
F:2405-2577/Domain: von Willebrand factor type A repeat homology <VM11>
F:2623-2806/Domain: von Willebrand factor type A repeat homology <VM12>
F:2803-2846/Domain: platelet glycoprotein Ib-like #status predicted <GPI>
F:2954-3039/Domain: fibronectin type III repeat homology <FN3>
F:3072-3132/Domain: animal Knittz-type proteinase inhibitor homology <BPI>
F:201,2084,2436,2563,2581,2683,2867,2920,3003/Binding site: cardonhydrate (Asn) (covalent

Alignment Scores:

Pred. No.:	4,86e-06	Length:	3137
Score:	230.50	Matches:	76
Percent Similarity:	47.1%	Conservative:	31
Best Local Similarity:	33.5%	Mismatches:	105
Query Match:	9.7%	Indels:	15
DB:	2	Gaps:	7

US-10-699-035a-5 (1-1254) x A37797 (1-3137)

```

QY      4 CTCGCGCTGGACGCGG--CTCGGCTGGCGCTTGAAGCTTGGCGGCTGGCGGAGC 60
      |||||
DB      7 LeuProLeuAlaLeuLeuGlyLeuLeuLeu-----GlyPheCysSerVal 23
      |||||
QY      61 GGGCGGAGCGCGGTCACAGCATGACCCCGA-----GGGAGCTGATG 108
      |||||
DB      24 GlyAlaGlnGln-----GlnAlaAlaValArgAsnValAlaValAlaAspIleIle 40
      |||||
QY      109 TTCCTGCTGGACAGCTCAGCAGCGTCTCTACATGAGTTCTCCGGGTTCCGAGATT 168
      |||||
DB      41 PheLeuValAspSerSerTrpSerIleGlyValGlnHisPheGlnLeuValArgGlnPhe 60
      |||||
QY      169 GTGGGGCAGCTGTGGCTCCACTGCCCCCTGGGACACCGGGGCGCTGCGTCCAGTGTG 228
      |||||
DB      61 LeuTyrAspValValIysAlaLeuAspValGlyGlyAsnAspPheArgPheAlaLeuVal 80
      |||||
QY      229 CAGTGGGCGAGTGGCGCATACACGAGTTCCCTTCGGCCAGACAGCTGGGAGGCT 288
      |||||
DB      81 GlnPheSerGlyAsnProHisTrpGlnPheGlnLeuAsnThrTyrProSerAsnGlnAsp 100
      |||||
QY      289 GCCCAGGATGCGGTGCGTCTTCCGACGCGATGGGTGACACCACTGGCGCTGGCG 348
      |||||
DB      101 ValLeuSerHisIleLeuAlaAsnMetProTyrMetGlyGlySerIleThrGlyValGly 120
      |||||
QY      349 CTGGCTTATGCAAGAAAGCTGTTTCTGAAGCATCAGTGCCTCCG-----CCAGGG 402
      |||||
DB      121 LeuGluTyrLeuIleGluAsnHisLeuThrIleAlaGlySerArgAlaSerGluGly 140
      |||||
QY      403 GTGCCCAAAAGTGTGGTGGTGAAGATGGGGGCTCCGACGACCTGTGGGCCCCC 462
      |||||
DB      141 ValProGlnValIleIleValLeuThrAspGlyGlnSerGlnAspAspValAlaLeuPro 160
      |||||
QY      463 ATGACGAGAGCTCAAGACCTGGGCGCTTCATGTTCTTACAGACCGGCGAGGCAAC 522
      |||||
DB      161 SerSerValLeuIleValSerAlaHisValAsnMetIleAlaValGlyValGlnAspAlaVal 180
      |||||
QY      523 TTCCTGAGCTGTGACGCGCTGCTCAGCCCTGCGGAGAGCACTG---CACTTTGTG 579

```

```

DB      181 GluGlyGlnLeuIleGluIleAlaSerArgProPheAspThrHisLeuPheAsnLeuGlu 200
      |||||
QY      580 GACGTGATGACCTGCACATCAATTGCAAGAGCTGAGGGGCTGCATT---CTCGGAGT 636
      |||||
DB      201 AsnPheThrAlaLeuHisGlyIleValGlyAspLeuValAlaSerValArgThrSerMet 220
      |||||
QY      637 CGGCGCAGCAGCTCCATGCC 657
      |||||
DB      221 ThrProGluGlnAlaGlyAla 227
      |||||

```

Search completed: February 13, 2006, 13:41:49
Job time : 78.062 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:13:13 ; Search time 63.7853 Seconds
(without alignments)
2774.097 Million cell updates/sec

Title: US-10-699-035A-5

Perfect score: 2380
Sequence: 1 atgcctccctgcagcgct.....ccgcacgcgcgtgcgcgttaa 1254

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 216443 segs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xid
-DB=/abs/ABSSREF.spool/US10699035/runat_13022006_062441_25416/app.query.fasta_1
-DB=uniprot -QFMT=fasta -SUFFIX=rup -MIMATCH=0.1 -LOOFCU=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcp -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=US10699035 @CGN 1.1 580 @runat_13022006_062441_25416 -NCP=6 -ICPU=3
-NO MAMP -NEG SCORES=0 -WAIT -DSPBLLOC=100 -LONGLOG -DEV TIMEOUT=120
-NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Uniprot 05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2114	88.8	445	2	06PCB0_HUMAN	06PCB0 homo sapien
2	1644.5	69.1	415	2	08R225_MOUSE	08R225 mus musculu
3	1640.5	68.9	415	2	0923K3_MOUSE	0923K3 mus musculu
4	1640.5	68.9	415	2	08C0Q7_MOUSE	08C0Q7 mus musculu
5	1627	68.4	415	2	0642A6_RAT	0642A6 rattus norv
6	1279	53.7	281	2	0715D7_HUMAN	0715D7 homo sapien
7	1058.5	44.5	233	2	09H6J5_HUMAN	09H6J5 homo sapien
8	786	33.0	203	2	08DVY9_MOUSE	08DVY9 mus musculu
9	672.5	28.3	505	2	04SCD1_TETNG	04SCD1 tetradon n
10	421.5	18.7	3124	2	05YVK2_HUMAN	05YVK2 gallus gall
11	421.5	17.7	2884	2	05YVK2_HUMAN	05YVK2 homo sapien
12	421.5	17.6	3063	2	05YVK1_HUMAN	05YVK1 homo sapien
13	419	17.6	3063	1	05YVK1_HUMAN	05YVK1 homo sapien
14	411.5	17.3	1723	2	04SD22_TETNG	04SD22 tetradon n
15	404.5	17.0	3119	1	04C0A1_MOUSE	04C0A1 mus musculu
16	398.5	16.7	517	2	043853_HUMAN	043853 homo sapien

17	393	16.5	1117	2	04RXN8_TETNG	04RXN8 tetradon n
18	392	16.5	1259	2	04RP12_TETNG	04RP12 tetradon n
19	389	16.3	1297	2	04VXQ4_HUMAN	04VXQ4 homo sapien
20	389	16.3	1329	1	K1510_HUMAN	09P218 homo sapien
21	388	16.3	1284	2	04VXQ5_HUMAN	04VXQ5 homo sapien
22	388	16.3	1284	2	06P159_HUMAN	06P159 homo sapien
23	387.5	16.3	637	2	081VX1_HUMAN	081VX1 homo sapien
24	387	16.3	2944	1	0707A1_HUMAN	0707A1 homo sapien
25	386.5	16.2	1253	1	097566_CANPA	097566 canis fami
26	382	16.1	1888	1	06EAL1_CHICK	P32018 gallus gall
27	379.5	15.9	2936	2	07YRK8_CANPA	07YRK8 canis fami
28	374.5	15.7	2225	2	04SKX3_TETNG	04SKX3 tetradon n
29	361.5	15.2	2944	2	063870_MOUSE	063870 mus musculu
30	355.5	14.9	1472	2	0902A0_CHICK	0902A0 gallus gall
31	341.5	14.3	1797	1	06EAL1_MOUSE	080X19 mus musculu
32	324	13.6	1557	2	04SH63_TETNG	04SH63 tetradon n
33	323.5	13.6	1796	1	06EAL1_HUMAN	005707 homo sapien
34	322	13.5	534	2	04GOW1_HUMAN	04GOW1 homo sapien
35	317	13.3	1816	2	06NT15_BRABE	06NT15 brachydanio
36	311	13.1	695	2	04NX38_9DELT	04NX38 anaeromyxob
37	310.5	13.0	1060	2	04NRA5_9DELT	04NRA5 anaeromyxob
38	291	12.2	2268	2	04SO43_TETNG	04SO43 tetradon n
39	287	12.1	828	2	04NOY7_9DELT	04NOY7 anaeromyxob
40	287	12.1	857	2	04NRP2_9DELT	04NRP2 anaeromyxob
41	285.5	12.0	2147	2	04NUV3_9DELT	04NUV3 anaeromyxob
42	283	11.9	964	2	04NOB4_9DELT	04NOB4 anaeromyxob
43	283	11.9	1549	2	04NU10_9DELT	04NU10 anaeromyxob
44	282.5	11.9	608	2	0905K9_CHV12	0905K9 cercopithec
45	282.5	11.9	2268	2	04NRA5_9DELT	04NRA5 anaeromyxob

ALIGNMENTS

RESULT 1
06PCB0_HUMAN
ID 06PCB0_HUMAN PRELIMINARY; PRT; 445 AA.
AC 06PCB0;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein, isoform 1 (WARP).
GN Name=WARP; ORFNames=RP4-758J18.11-001;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.J.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Raskar S., Weisley K.C., Hale S., Garcia A.M., Gay L.J., Huiy S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Feilley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Scherf A., Schen J.E., Krzywicki M.I., Skalska U., Smilans D.E.,
RA Scherf A., Schen J.E., Jones S.J.M., Maria M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.

RA Strausberg R.D., Collins E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemkin C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheefter C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gumarate P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
RA Paley J., Hailton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Mhling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz E., Myers R.M.,
RA Buttefield J.W., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=Czech II;
RC TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RG NIH MCC Project;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC MEDLINE=1085660; PubMed=1121751; DOI=10.1038/35055500;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.,
RT "High-efficiency full-length cDNA cloning",
RL Meth. Enzymol. 303:19-44(1999).
RN [4]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC MEDLINE=1085660; PubMed=1121751; DOI=10.1038/35055500;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Iehi Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kodera K., Matsuda H.A., Ashburner M.; Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Knehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Maehina T., Mazzarelli J., Mombaeets P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker Y., Wilting L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kottuk S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.",
RL Nature 409:685-690(2001).
RN [5]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RL Nature 420:563-573(2002).
RN [6]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RX Carninci P., Shibata Y., Hayatsu N., Sugihara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.",

Genome Res. 10:1617-1630(2000).

[7] NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=20530913; Pubmed=11076861; DOI=10.1101/gr.152600;
RA Shihata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishie T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:11757-1771(2000).

[8] NUCLEOTIDE SEQUENCE.

RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirakawa T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Koyama M., Koya S.,
RA Kuribara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shihata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

[9] NUCLEOTIDE SEQUENCE.

RP STRAIN=FVB/N; TISSUE=Colon;
RC Director MGC Project;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC026919; AAH26919.1; -; mRNA.
DR EMBL: AK077240; BAC36703.1; -; mRNA.
DR EMBL: BC061166; AAH36166.1; -; mRNA.
DR HSP: P18614; IMHP.
DR Ensembl: ENSMUSG0000042116; Mus musculus.
DR MGI: MGI:2179729; W001.
DR GO: GO:0005615; Cytoextracellular space; TAS.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR002035; VWF_A.
DR Pfam: PF00041; Fn3_2.
DR Pfam: PF00092; VWA_1.
DR PRINTS: PR00453; VMPADOMAIN.
DR SMART: SM00060; FN3_2.
DR SMART: SM00327; VWA_1.
DR PROSITE: PS50853; FN3_2.
DR PROSITE: PS50234; VWF_A_1.
SQ SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;

Alignment Scores:

Pred. No.:	2,38e-79	Length:	415
Score:	1644.50	Matches:	325
Percent Similarity:	85.9%	Conservative:	33
Best local Similarity:	77.9%	Mismatch:	56
Query Match:	69.1%	Indels:	3
DB:	2	Gaps:	1

US-10-699-035A-5 (1-1254) x OBR225_MOUSE (1-415)

OY 1 ATGCTCCCTGGAGCGCGCTGGCGCCGAGCTTGCGGCTGCGCGCGAGC 60

Db 1 MetLeuPheTrpThrAlaPheSerSerThrAlaLeuSerLeuAlaValaIaIaSer 20

OY 61 GGCAGCGAGCGCGCTGCACACGATCACGCCGCCGAGGAGGAGCTGATGTTCTGCTGAC 120

Db 21 SerTleGluArgGlySerThrAlaSerAspProGlnGlyAspLeuLeuPheLeuLeuAsp 40

OY 121 AGCTACGCAAGCGCTCTCTACTACAGAGTTTCCGCGGCTTGCGGAGTTTGAGGGCAGCTG 180

Db 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACGTCCTCGGACCGGGGCGCTGCGTGCACGTCGTGACGTCGGGCACT 240
Db 61 ValAlaThrMetSerPheGlyProGlyAlaLeuArgAlaSerLeuValHisIstValGlySer 80
QY 241 CGGGCATACACGAGTTCCTCCGCGACACAGTCGGGGTGAAGCTTCCACAGATGG 300
Db 81 GlnProHisIstHrGluPheHrPheAspGlnTyrSerSerGlyGlnAlaIleGlnAspAla 100
QY 301 GTGGGCTTCGTCGCGGACGATGGGTGACACCCACACTGTCGGCTGCGGCTGTGCG 360
Db 101 IleArgValAlaProGlnHrMetGlyAspThrAsnThrGlyLeuAlaLeuAlaTyrAla 120
QY 361 AAGGACACAGCTGTTTGTCTGAAGCATGAGTCCCGGCGACGGGTGCGCAAGTGTGTG 420
Db 121 LysGluGlnLeuPheHrIleGluGlnAlaGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTGACAGATGGCGGCTTCACGACCTGTGGGCGCGCGCGCGCGCGCGCGCGCG 480
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProProMetGlnGluLeuLysAsp 160
QY 481 GTGGGCTGACCGTGTTCATTGTCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 161 LeuGlyValThrIlePheIleValSerThrGlyArgGlyAsnLeuLeuGluLeuLeuAla 180
QY 541 GGTGCTCCACGTCCTCGGACCGGACGACCTGACTTGTGAGCTGATGACCTGCACATC 600
Db 181 AlaAlaSerAlaProIleGluGlySerHisIlePheValAspValAspAspLeuProIle 200
QY 601 ATTGTCCAAAGAGTGAAGGAGCTTCATTCTC--GCGATGGCGCGCGACGACTCCATGCC 657
Db 201 IleAlaArgGluLeuArgGlySerIleThrAspAlaMetGlnProGlnGlnLeuHisAla 220
QY 658 ACGGAGTACCGTCGCGGCTTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Db 221 SerGluValLeuSerSerGlyPheAspGluLeuSerTrpProLeuLeuThrAlaAspSer 240
QY 718 GGCTACTATGTGTGAGTGTGTGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
Db 241 GlyTyrTyrValLeuGlnLeuValProSerGlyLysLeuAlaThrThrArgGlnGln 260
QY 778 CTGGCCAGGGAACCGCCACGACTGGATCTGGGCGCGCGCTGACCGGACACGACTGCAC 837
Db 261 LeuProGlyAsnAlaThrSerTrpThrTrpThrAspLeuAspProAspThrAspTyrGln 280
QY 838 GTGGCGCTGAGTGTGATGTCACAGTCGCGCTCTGAGGCGCGCGAGATCCGCGGCGCG 897
Db 281 ValSerLeuLeuProGlnSerAsnValHisLeuLeuArgProGlnHisValArgValArg 300
QY 898 ACGCGGCGACGAGGCGCGGCGCGACGCGATCGTATCTCCACGCGCGCGCGCGCGCAG 957
Db 301 ThrLeuGlnGlnGlnAlaGlyProGlnLysGlnIleValIleSerHisAlaArgProArgSer 320
QY 958 CTTCGCGTGAATTGGGCGCGCGCGCTGACGCGCGCGCGCTGCGGCTTACACGTCGAC 1017
Db 321 LeuArgValSerTrpAlaProAlaLeuGlyProAspSerAlaLeuGlyTyrHisValGln 340
QY 1018 TTTCGGGCGCTGCGGCGGCGGAGCGGCGCGGCGGAGTGTGCGCGCGCGCGCACTGC 1077
Db 341 LeuGlyProLeuGlnGlyGlySerLeuGlnArgValGlnValProAlaGlyLysAsnSer 360
QY 1078 ACCACCTGACGAGGCTGCGCGCGGCGCGCGCTACCTGATCGTGCACGTCGCGCGCTTC 1137
Db 361 ThrThrValGlnGlnLysLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe 380
QY 1138 CGCTGGGCGCGGACGCGCGCTGTCCGCGACGCTGACGCGCGCGCGCGCGCGCGCG 1197
Db 381 ArgSerGlyArgGlnArgAlaLeuSerAlaLysAlaCysThrAlaSerGlyAlaArgThr 400
QY 1198 CGCGCACG 1244
Db 401 Arg-AlaProGlnSerMetArgProGlnAlaGlyProArgGlnPro 415

RESULT 3
Q923K3_MOUSE
ID Q923K3_MOUSE PRELIMINARY; PRT; 415 AA.
AC Q923K3.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Von Willebrand factor A-related protein.
GN Name=Wval; Synonyms=4932416A1Rik, Wap;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC NCBI_Taxid=10090;
OX NCBI_Taxid=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22057805; PubMed=12062410; DOI=10.1016/S0014-5793(02)02579-6;
RA FitzGerald J., Ting S.T., Bateman J.F.;
RT "WAB a new member of the von Willebrand factor A-domain superfamily
of extracellular matrix proteins.";
RL FEBS Lett. 517:61-66(2002).
DR EMBL; AY030094; AAK38350.1; -; mRNA.
DR HSSP; P18614; 1MRP.
DR Ensembl; ENSMUSG00000042116; Mus musculus.
DR MGI; MGI:2179729; 4932416A1Rik.
DR MGI; MGI:2179729; Wval.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002035; vWF_A.
DR Pfam; PF00041; FN3; 2.
DR Pfam; PF00092; vWA; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00327; vWA; 1.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS50234; vWFA; 1.
SQ SEQUENCE 415 AA; 44737 MW; C3344AECB3FDE431 CRC64;

Alignment Scores:
Pred. No.: 3.88e-79 Length: 415
Score: 1640.50 Matches: 324
Percent Similarity: 85.9% Conservative: 34
Best Local Similarity: 77.7% Mismatch: 56
Query Match: 68.9% Indels: 3
DB: Gaps: 1

US-10-699-035A-5 (1-1254) x Q923K3_MOUSE (1-415)
QY 1 ATGCTCCCTGGACGCGCGCTCGGCTGCGCTTGCAGCTTGCAGCTGCGCGCGGACG 60
Db 1 MetLeuPheTrpThrAlaPheSerMetAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
QY 61 GCGCGGAGCGCGGCTTCACACGACATCAGCCCGCGGAGGAGACATGCTTCTGTCGAC 120
Db 21 SerIleGluArgGlySerThrThrAlaSerAspProGlnGlyAspLeuLeuPheLeuLeuAsp 40
QY 121 AGCTGACGCGGCTCTCTCCTACAGAGTTCCTCCGGTTCGGAGTTCGTGGCGAGCTG 180
Db 41 SerSerAlaSerValSerHisIstHrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACGTCCTCGGACCGGGGCGCTGCGTGCACGTCGTGACGTCGGGCACT 240
Db 61 ValAlaThrMetSerPheGlyProGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGGCATACACGAGTTCCTCCGCGACACAGTCGGGGTGAAGCTTCCACAGATGG 300
Db 81 GlnProHisIstHrGluPheHrPheAspGlnTyrSerSerGlyGlnAlaIleGlnAspAla 100
QY 301 GTGGGCTTCGTCGCGGACGATGGGTGACACCCACACTGTCGGCTGCGGCTGTGCG 360
Db 101 IleArgValAlaProGlnHrMetGlyAspThrAsnThrGlyLeuAlaLeuAlaTyrAla 120

QY 361 AAGGACAGCTGTTGTCAGACATCAGGTGCCCCGAGGGTCCCAAGTGTG 420
DB 121 LysGlnGlnLeuPheHlaGlnGlnAlaArgProGlnValProLysValLeuVal 140
QY 421 TGGGTCAGAGTGGGGCTCCAGGACCCCTGGGGCCCCCGCCAGGAGCTCAAGGAC 480
DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheGlnGlnLeuLysAsp 160
QY 481 CTGGGGCTCAGCTGTTTCATGTCAGACGCGCCGAGGCAACTTCCTGAGCTGTCAGCC 540
DB 161 LeuGlyValThrLeuPheHlaValSerThrGlyArgGlyLysLeuLeuGlnLeuLeuAla 180
QY 541 GCTGCTCAGCCCTGCGGAGAACACCTGACTTGTGACGTGATGACTTCGACATC 600
DB 181 AlaAlaSerAlaProAlaGlnGlyShiLeuHlaPheValAspValAspLeuProIle 200
QY 601 ATTGTCCAGAGTGGAGGCTCCATTCCTC--CGGATGGGGCGGAGGAGCTCCATGCC 657
DB 201 LLeAlaArgGlnLeuLysGlySerLLeThrAspAlaMetGlnProGlnGlnLeuHlaAla 220
QY 658 ACCGAGATCAGCTCCAGCGGCTCCGCTGAGCTGAGCCAGCTGTCAGCGAGATCG 717
DB 221 SerGlnValLeuSerSerGlyPheAspArgLeuSerTrpProLeuLeuThrAlaAspSer 240
QY 718 GGGTACTATGTGCTGAGACTGTGCCCCAGCGCCAGCGGGGCTGACAGAGCCAGAG 777
DB 241 GlyTyrTyrValLeuGlnLeuValProSerGlyLysLeuAlaThrThrArgGlnGln 260
QY 778 CTGGCAGGAGAACCGCAGCAGCTGAGTCTGGGCGGCTGACCCGAGACAGGACTAGCAG 837
DB 261 LeuProGlyAspAlaThrSerTrpThrTrpThrAspLeuAspProAspThrAspTyrGln 280
QY 838 GTGGCCCTAGTGCCTGAGTCCAGCTGAGCTGAGCGCCAGCTCCAGTCTCCGAGGAGCGC 897
DB 281 ValSerLeuLeuProGlnSerAlaHlaLeuLeuArgProGlnHlaValAlaArgValArg 300
QY 898 ACCGCGCAGAGAGGCGCGGCGCAGAGCGGATGTCATCTCCAGCGCCGCGCGCAGC 957
DB 301 ThrLeuGlnGlnGlnAlaGlyProGlnArgGlnValIleSerHlaAlaArgProArgSer 320
QY 958 CTCCGCGTGAAGTGGGCGCCAGCGGCTGAGGCTCAGCGCGGCTCGGCTACACAGTCGAG 1017
DB 321 LeuArgValSerTrpAlaProAlaLeuGlyProAspSerAlaLeuGlyTyrHlaValGln 340
QY 1018 TTGCGGCGCGCTGCGGCGGAGGCGCAGCGGCTGAGTGGCGCGCGCGCCAGCATCG 1077
DB 341 LeuGlnProLeuGlnGlnGlySerLeuGlnArgValGlnValProAlaGlyGlnAspSer 360
QY 1078 ACCAGCCTGACAGGCGCTGCGCGCGGCGCAGCGCTTACCTGTGACCGTGAACCGCGCTTC 1137
DB 361 ThrThrValGlnGlnLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe 380
QY 1138 CGCTCGGCGCGCAGAGCGGCTGTGCGGCAAGGCTTCGACGCGCCAGCGCGCGCGCG 1197
DB 381 ArgSerGlyArgGlnArgAlaLeuSerAlaValAlaCysThrAlaSerGlyAlaArgThr 400
QY 1198 CGCGCAGCGCGCGCGCGCGCGCGCGCGAGCGGAGCGCGCGCGCG 1244
DB 401 Arg--AlaProGlnSerMetArgProGlnAlaGlyProAlaGlnPro 415

RESULT 4
Q8C007 MOUSE PRELIMINARY; PRT; 415 AA.
AC Q8C007
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932416A11 product:VON WILLEBRAND FACTOR A-RELATED PROTEIN homolog
GN Name=Wval; Synonyms=4932416A11Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]:
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.,
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=2030913; PubMed=11076661; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Katsunari T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., Mowman P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
 RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marx M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=pancreas;
 RG NIH MGC Project;
 RU Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC005543; AM03543.2; -, mRNA.
 DR InterPro: IPR003961; FN_III.
 DR Pfam: PF00041; fn3; 2.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PS50853; FN3; 2.
 FT NON_TER
 SQ SEQUENCE 281 AA; 29628 MW; 350CCBA590791BD4 CRC64;

Alignment Scores:
 Pred. No.: 5, 89e-60 Length: 281
 Score: 1279.00 Matches: 253
 Percent Similarity: 90.0% Conservative: 0
 Best Local Similarity: 90.0% Mismatches: 0
 Query Match: 53.7% Indels: 28
 DB: 2 Gaps: 2

US-10-699-035A-5 (1-1254) x Q7L5D7_HUMAN (1-281)

QY 493 GTGTCATTGTGACGACCGGCGAGGCACTTCCTGAGCTGACCGCTGCTGACGC 552
 Db 1 ValPheIleValSerThrgIyArgIyAAsnPheIleuGluLeuSerIalaIaIaSerAla 20
 QY 553 CCGCGGAGACGACCTGACCTTGTGACGTGGAGACCTGACATCTGTCCTGACGAG 612
 Db 21 ProAlaGluIuYshIleuHIsPheValaAspValaAspPheuHIsIleIleValGInu 40
 QY 613 CTGAGGGGCTTCATTCTC---GCGATCGGCGCGACGACCTCCATGCCAGAGATCAG 669
 Db 41 LeuArgGlySerIleuAspIaIaMetArgProGInGInIeHIsIaIaThrgIuIeThr 60
 QY 670 TCCAGGGCTTCGCGCTGCGCTGCGCAACCCCTGCTGACCGGACGACTCGGGCTAATGTG 729
 Db 61 SerSerGlyPheArgIeuaIaIaTrpProIeuleuThraIaAspSerGlyTyTyVal 80
 QY 730 CTGGAGCTGGTGGCCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGC 789
 Db 81 LeuGluLeuValProSerIaIaGInProGlyAlaIaIaArgIuInIeupProGlyAAsn 100
 QY 790 GCCACGAGCTGATCTGGGCGCGCTCGACCCCGGACGAGACTCCAGCTGGCGCTAATG 849
 Db 101 AlaThraAspTrpIleThraIaGlyLeuAspProAspThraAspIyAspValaIaIeVal 120
 QY 850 CCTGAGTCCAGCTGGCGCTTCCTGAGGCGCCGAGACTCTGCGGGTGGCGCGG----- 903
 Db 121 ProGluSerAsnValaIaIaGluLeuArgProGInIleIeuaIaIaIaThraIaIaIa 140
 QY 903 ----- 903
 Db 141 GluAlaGlyProGlyAlaSerGlyProGluSerGlyAlaGlyProAlaProThrgInIe 160
 QY 904 -----CCAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 948
 Db 161 AlaAlaIeupProAlaIaIaProGluGluAlaGlyProGluArgIleValIleSerHIsaIaIa 180
 QY 949 CCGCGGAGCTCGCGCTGAGTTGGGCCCGGCGGCTGAGCGCGGCGGCTCGGCTAC 1008

Db 161 ProArgSerIeuaIaIaValSerTrpAlaProAlaIeugIySerAlaIaIaIeugIyTy 200
 QY 1009 CACGTGCAATTCGGGCGCGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGG 1068
 Db 201 HisValGInPheGlyProIeuaIaIaGlyIyGlyAlaIaIaIaIaIaIaIaIaIaIa 220
 QY 1069 CGCAATCGACACGCTGACGAGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1128
 Db 221 ArgAsnCysTrpThraIeugInGlyIeuaIaIaProGlyThraIaIaIaIaIaIaIaIa 240
 QY 1129 GCGGCTTCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1188
 Db 241 AlaAlaIa 260
 QY 1189 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1248
 Db 261 ProArgProArgProArgProArgProArgProArgProArgProArgProArgProArg 280
 QY 1249 CCG 1251
 Db 281 Pro 281

RESULT 7
 Q9H6J5_HUMAN
 ID Q9H6J5_HUMAN PRELIMINARY; PRT; 233 AA.
 AC Q9H6J5;
 DT 01-MAR-2001 (TRENBLREL. 16, Created)
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
 DE Hypothetical protein FLN22215.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
 OC Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 RA Suzuki Y., Ohayashi M., Nishii T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isgai T., Sugano S.;
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AK025868; BAB15264.1; -, mRNA.
 DR HSP; Q96KP7; 1FNA.
 DR Ensembl; ENSG00000179403; Homo sapiens.
 DR InterPro: IPR003961; FN_III.
 DR Pfam: PF00041; fn3; 2.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PS50853; FN3; 2.
 SQ SEQUENCE 233 AA; 24473 MW; B2CC118FC5B7BC50 CRC64;

Alignment Scores:
 Pred. No.: 2, 95e-48 Length: 233
 Score: 1058.50 Matches: 206
 Percent Similarity: 88.4% Conservative: 0
 Best Local Similarity: 88.4% Mismatches: 0
 Query Match: 44.5% Indels: 27
 DB: 2 Gaps: 1

US-10-699-035A-5 (1-1254) x Q9H6J5_HUMAN (1-233)

QY 634 ATGCGGCGGCGGAGCTCCATGCCAGAGATCACGTCGAGGCTTCGCGCTGCGCTGG 693
 Db 1 MetArgProGInGInIeHIsIaIaThrgIuIeThraSerSerGlyPheArgIeuaIaIa 20
 QY 694 CCAACCTCGTGAACCGGACGCTGAGCTGCTAATGTGCTGAGCTGAGCTGCGGCGGCGG 753
 Db 21 ProProIeuleuThraIaAspSerGlyTyTyValIeugIuLeuValProSerIaIaGIn 40
 QY 754 CCGGGGGCTGCAAGACCGGACGAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 813
 Db 41 ProGlyAlaIa 60

QY 814 CTGAGCCGAGACAGGACTACAGCTGGCGCTAGTGCCTGAGTCCAGGCGCTCTG 873
DB 61 LeuaspProaspThraspTyraValAlaLeuValProGluSerAsnValArgLeuLeu 80
QY 874 AGGCGCCAGATCTGTGGGGTGGCAGCGG----- 903
DB 81 ArgProGlnHleuValArgValArgThrArgProGlyGluAlaGlyProGlyAlaSerGly 100
QY 904 -----CCAGAGAG 912
DB 101 ProGluSerGlyAlaGlyProAlaProThrGlnLeuAlaLeuProAlaProGluGln 120
QY 913 GCCGCGCCAGAGCGCATCTGATCTCCACGCCCGCGCGCAGCTTCGCGTGAAGTGG 972
DB 121 AlaGlyProGluArgGlyLeuValIleSerHisAlaArgProArgSerLeuArgValSerTyr 140
QY 973 GCCCGCGCGCTGGGCTCAGCGCGCGCTGGCTACCACTGAGTTCGGCGCGCTGGCG 1032
DB 141 AlaProAlaLeuGlySerAlaAlaAlaLeuGlyTyrHisValGlnPheGlyProLeuArg 160
QY 1033 GCGCGGAGCGCGAGCGGCTGAGTGCCTCCGCGCGCGCAGCTGACAGCGCGC 1092
DB 161 GlyGlyGluAlaGlnArgValGluValProAlaGlyArgAsnGlyThrThrLeuGlnGly 180
QY 1093 CTGGCGCCGCGGACCGGCTTACTGTGAGCGGTGACCGCGCGCTTCGCTCGCGCGCGAG 1152
DB 181 LeuAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPheArgSerGlyArgGlu 200
QY 1153 AGCGCGCTGCGCGCAGCGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1212
DB 201 SerAlaLeuSerAlaLysAlaLysAlaCysThrProAlaGlyProArgProArgProAla 220
QY 1213 CCGCGCGCGCGCGAGCGCGCGGAGCGCGAGCGCGTGAAGCGG 1251
DB 221 ProArgAlaProThrProGlyThrAlaSerArgGluPro 233

RESULT 8
Q8VDV9_MOUSE PRELIMINARY; PRT; 203 AA.
ID Q8VDV9_MOUSE PRELIMINARY; PRT; 203 AA.
AC Q8VDV9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 4932416A1Rik protein.
GN Name=Vwal; Synonyms=4932416A1Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX MEDLINE=23288257; PubMed=12477932; DOI=10.1073/pnas.242608999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schlier G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gamaralline P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Schermer A., Schein J.E., Jones S.J.W., Maitra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RA Strausberg R.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020136; AAH20136.1; -, mRNA.
DR Ensembl; ENSMUSG00000042116; Mus musculus.
DR MGI; MGI:2179729; 4932416A1Rik.
DR MGI; MGI:2179729; Vwal.
DR GO; GO:0005615; Cytoplasmic space; TAS.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00853; FN3; 2.
KW Repeat.
SQ SEQUENCE 203 AA; 22068 MW; AAD7FDEF37626898 CRC64;

Alignment Scores:
Pred. No.: 8, 52e-34 Length: 203
Score: 786.00 Matches: 152
Percent Similarity: 83.4% Conservative: 19
Best Local Similarity: 74.1% Mismatches: 32
Query Match: 33.0% Indels: 2
DB: Gaps: 0

US-10-699-035A-5 (1-1254) x Q8VDV9_MOUSE (1-203)

QY 634 ATGCGCGCGCAGACAGCTCCATCCACGAGATCACGCTCCAGCGCTTCGCGCTGCG 693
DB 1 MetGlnProGlnGlnLeuHisAlaSerGluValLeuSerSerGlyPheArgLeuSerTyr 20
QY 694 CCACCCCTGTCAGCGCAGACACTCGGCTACTATGCTGCTGAGCTGCGCGCGCGCGCG 753
DB 21 ProProLeuLeuThrAlaAspSerGlyTyrValLeuGlnLeuValProSerGlyLys 40
QY 754 CCGGCGGCTGCAAGACGCGCAGCAGCTGCGCAGGAAAGCCACGAGTGTGAGCGCGCG 813
DB 41 LeuAlaThrThrArgArgGlnGlnLeuProGlyAsnAlaThrSerTyrThrThrAsp 60
QY 814 CTGAGCCGAGACAGGACTACAGCTGGCGCTAGTGCCTGAGTCCAGGCGCTCTG 873
DB 61 LeuaspProaspThraspTyraValAlaLeuValProGluSerAsnValArgLeuLeu 80
QY 874 AGGCGCCAGATCTGTGGGGTGGCAGCGG----- 903
DB 81 ArgProGlnHleuValArgValArgThrArgProGlyGluAlaGlyProGlyAlaSerGly 100
QY 904 -----CCAGAGAG 912
DB 101 ProGluSerGlyAlaGlyProAlaProThrGlnLeuAlaLeuProAlaProGluGln 120
QY 913 GCCGCGCCAGAGCGCATCTGATCTCCACGCCCGCGCGCAGCTTCGCGTGAAGTGG 972
DB 121 AlaGlyProGluArgGlyLeuValIleSerHisAlaArgProArgSerLeuArgValSerTyr 140
QY 973 GCCCGCGCGCTGGGCTCAGCGCGCGCTGGCTACCACTGAGTTCGGCGCGCTGGCG 1032
DB 141 AlaProAlaLeuGlySerAlaAlaAlaLeuGlyTyrHisValGlnPheGlyProLeuArg 160
QY 1033 GCGCGGAGCGCGAGCGGCTGAGTGCCTCCGCGCGCGCAGCTGACAGCGCGC 1092
DB 161 GlyGlyGluAlaGlnArgValGluValProAlaGlyArgAsnGlyThrThrLeuGlnGly 180
QY 1093 CTGGCGCCGCGGACCGGCTTACTGTGAGCGGTGACCGCGCGCTTCGCTCGCGCGAG 1152
DB 181 LeuAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPheArgSerGlyArgGlu 200
QY 1153 AGCGCGCTGCGCGCAGCGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1212
DB 201 SerAlaLeuSerAlaLysAlaLysAlaCysThrProAlaGlyProArgProArgProAla 220
QY 1213 CCGCGCGCGCGCGAGCGCGCGGAGCGCGAGCGCGTGAAGCGG 1251
DB 221 ProArgAlaProThrProGlyThrAlaSerArgGluPro 233

QY 934 ATCTCCACAGCGCGCGCGCGCGCTCGCGTGAAGTGGAGCGCGCTGAGCGCTCAGCG 993
DB 101 IleSerHisAlaArgProArgSerLeuArgValSerTyrAlaProAlaLeuGlyProAla 120
QY 994 GCGCGCGCTGCGTACCACTGACAGTTCGCGCGCTGCGCGCGCGCGCGCGCGCGG 1053
DB 121 SerAlaLeuGlyTyrHisValGlnLeuGlyProLeuGlnGlyGlySerLeuGlnArgVal 140
QY 1054 GAGGTGCGCGCGCGCGCGCACTGACAGCGCTGCGAGGCGCTGCGCGCGCGCGCTAC 1113
DB 141 GluValProAlaGlyGlnIleSerHisThrThrValGlnGlyLeuThrProCysThrThrTyr 160
QY 1114 CTGAGTACCGTACAGCGCGCTTCGCTGAGGCGCGGAGAGCGCGCTGCGCGCAGGCG 1173
DB 161 LeuValThrValThrAlaAlaPheArgSerGlyArgGlnArgAlaLeuSerAlaLys 180
QY 1174 TGCAGCGCGCAGCGG 1233
DB 181 CysThrAlaSerGlyAlaArgThrArg--AlaProGlnSerMetArgProGluAlaGlyP 200
QY 1234 ACCGCGAGCGG 1244
DB 200 roArgGluPro 203


```

FT CARBOHYD 890 890 O-linked (Xyl. . .) (chondroitin sulfate)
FT FT CARBOHYD 981 981 O-linked (Xyl. . .) (chondroitin sulfate)
FT FT CARBOHYD 1006 1006 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 1032 1032 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 1044 1044 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 1512 1512 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 1767 1767 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 2210 2210 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 2273 2273 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 2532 2532 N-linked (GlcNAc. . .) (Potential)
FT FT CARBOHYD 2683 2683 N-linked (GlcNAc. . .) (Potential)
FT FT VARSPLIC 25 1188 Missing (in isoform Short).
FT FT CONFLICT 1258 1258 /FTid=VSP_001148.
FT FT CONFLICT 1264 1264 T -> S (in Ref. 4).
FT FT CONFLICT 1264 1264 D -> E (in Ref. 4).
FT FT CONFLICT 2759 2759 P -> A (in Ref. 2).
FT FT CONFLICT 2803 2803 L -> F (in Ref. 2).
FT FT CONFLICT 2977 2977 V -> F (in Ref. 2).
FT FT CONFLICT 3075 3076 QP -> AG (in Ref. 3).
SQ SEQUENCE 3124 AA; 340582 MW; 094285AF7F346CF CRC64;

```

Alignment Scores:

```

Pred. No.: 1,23e-15 Length: 3124
Score: 444.00 Matches: 123
Percent Similarity: 47.2% Conservative: 64
Best Local Similarity: 31.1% Mismatches: 169
Query Match: 18.7% Indels: 40
DB: 1 Gaps: 7

```

US-10-699-035A-5 (1-1254) x COCAL_CHICK (1-3124)

```

QY 94 CGAGGAGACCTGATGTTCTGCTGACAGCTCAGCAGGCTCTCTACAGATTCTCC 153
   :: |||:::|||||::: ||| |||:::
DB 437 LysAlaAspValValPheLeuValLaepGlySerTyrSerIleGlyIleAlaAspIleVal 456
   ::|||:::|||||::: ||| |||:::
QY 154 CGAGTTCCGAGATTGTTGGGCGGAGCTGTTGCTTCCATCGCCCGGAGCGGCGCTG 213
   ::|||:::|||||::: ||| |||:::
DB 457 LysValAlaArgAlaPheLeuGluValLysSerPheGluIleSerProAlaIleVal 476
   ::|||:::|||||::: ||| |||:::
QY 214 CGTGCAGTCTGCTGACGTCGCGGACGTCGACATACCGGATCCCTCGGCGCCAGC 273
   ::|||:::|||||::: ||| |||:::
DB 477 GlnIleSerLeuValGlnTyrSerAlaAspProIleMetGlnSerLeuAsnArgIle 496
   ::|||:::|||||::: ||| |||:::
QY 274 AGCTCGGTGAGGCTGCCAGATGCGGTGCTTCTGCCAGCGCATGGGTGACACC 333
   ::|||:::|||||::: ||| |||:::
DB 497 AsnArgValLysAspIleIleGlnAlaIleAsnThrPheProTyrArgIleGlySerThr 516
   ::|||:::|||||::: ||| |||:::
QY 334 CACACTGGCCTGGCGCTGCTTATGCGCAAGAAACAGCTGTTGCTGACACATCAGTGC 393
   ::|||:::|||||::: ||| |||:::
DB 517 AsnThrGlyLysAlaMetThrTyrValArgGluValPheValThrSerIleGlySer 536
   ::|||:::|||||::: ||| |||:::
QY 394 CGGCGAGGAGTCCGCAAACTGCTGCTGCTGACAGTGGCGGCTCCGACCCCTG 453
   ::|||:::|||||::: ||| |||:::
DB 537 ArgProAsnValProArgValMetIleLeuIleThrAspGlyLysSerSerAspAlaPhe 556
   ::|||:::|||||::: ||| |||:::
QY 454 GGGCCCCCATGAGAGCTCAAGACCTGGCGCTCAGCGCTTCTGCTGACACCGGC 513
   ::|||:::|||||::: ||| |||:::
DB 557 LysGluProAlaIleLysLeuArgAspAlaAspValGlnIlePheAlaValGlyVal 576
   ::|||:::|||||::: ||| |||:::
QY 514 CGAGCAACTTCTGAGCTGTCAGCGGCTGCTGACCGCCCTGCGGAGACAGCCTGCAC 573
   ::|||:::|||||::: ||| |||:::
DB 577 AspAlaValArgThrGluLeuGluAlaIleAlaSerProProAlaGluThrHisValTyr 596
   ::|||:::|||||::: ||| |||:::
QY 574 TTTTGTG---GACCTGATGATCCTGCACATCATTTGTCACAGAGTGAAGGGCTCATCTTC 630
   ::|||:::|||||::: ||| |||:::
DB 597 ThrValGluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerVal 616
   ::|||:::|||||::: ||| |||:::
QY 631 GCGATCGGCGCGGAGAGCTCCATGCGC----- 657
   ::|||:::|||||::: ||| |||:::
DB 617 LeuArgIleGluGlnIleuAlaIleValIleArgLysSerTyrValProAlaLysAsn 636
   ::|||:::|||||::: ||| |||:::

```

```

QY 658 -----ACGAGATACAGCTCCAGCGGCTTCGCGCTGGCGCCACCCCTGCTGACC 708
   ::|||:::|||||::: ||| |||:::
DB 637 MetValPheSerAspValThrSerAspSerPheLysValSerTrpSerAlaIleGlySer 656
   ::|||:::|||||::: ||| |||:::
QY 709 GCAGACTCGGGCTACTATGTTGCTGAGCTGGGCGCCAGCGCCACCGGGGCTGCAGA 768
   ::|||:::|||||::: ||| |||:::
DB 657 GluGluLysSerTyrLeuIleLysTyrIleVal-----AlaIleGlyLysAspIlePhe 674
   ::|||:::|||||::: ||| |||:::
QY 769 CGCGAGAGCTCCAGGAGAACCCACAGCATGATCTGGGCGCGCTGACCCGAGACG 828
   ::|||:::|||||::: ||| |||:::
DB 675 IleValSerValProAlaSerSerThrSerSerValLeuThrAsnLeuLeuProIleThr 694
   ::|||:::|||||::: ||| |||:::
QY 829 GACTACGAGCTGCGCTAGTGCCTGATGCCATGCCAGTCCGCTCTGAGGCCACGATCTCG 888
   ::|||:::|||||::: ||| |||:::
DB 695 ThrTyrAlaValSerValIleAlaGlu-----Tyr 704
   ::|||:::|||||::: ||| |||:::
QY 889 CGGTCGCGACCGCGCCAGAGAGAGCGCGGCGGCGCATGTCATCTCCACGCGCG 948
   ::|||:::|||||::: ||| |||:::
DB 705 GluAspGlyAspGlyProProLeuAspGlyGluGluThrThrLeuGluValIleGlyAla 724
   ::|||:::|||||::: ||| |||:::
QY 949 CGCGCAGCTCCGCG-----GTGAGTTGGGCCCCA 978
   ::|||:::|||||::: ||| |||:::
DB 725 ProArgAsnLeuArgIleThrAspGluThrThrAspSerPheIleValGlyTrpThrPro 744
   ::|||:::|||||::: ||| |||:::
QY 979 GCGCTGGCTCAGCGCGCGCGCTCGGCTACCACTGCTGCGGCGCGCTGCGGCGCG 1038
   ::|||:::|||||::: ||| |||:::
DB 745 AlaProGlyAsn-----ValLeuArgTyrArgLeuValTyrArgProLeuThrIleGly 762
   ::|||:::|||||::: ||| |||:::
QY 1039 GAGCGGAGCGGAGTGGAGGTGCGCGGCGCGCACTGACACCGCTGAGGCGCTGCGG 1098
   ::|||:::|||||::: ||| |||:::
DB 763 GluArgArgGlnValThrValSerAlaAsnGluArgSerThrThrLeuArgAsnLeuIle 782
   ::|||:::|||||::: ||| |||:::
QY 1099 CGGAGCAGCGCTTACCTGCTGACCGTGAACCGCGCTTCGCTGCGGCGCGAGCGCG 1158
   ::|||:::|||||::: ||| |||:::
DB 783 ProAspThrArgTyrGluValSerValIleAlaGluTyrGlnSerGlyProGlyAsnAla 802
   ::|||:::|||||::: ||| |||:::
QY 1159 CTGTCCGCGAGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1206
   ::|||:::|||||::: ||| |||:::
DB 803 LeuAsnGlyTyrAlaLysThr---AspGluValArgGlyAsnProArg 817
   ::|||:::|||||::: ||| |||:::

```

RESULT 11

OSVYK2_HUMAN PRELIMINARY; PRT; 2884 AA.

```

ID OSVYK2;
AC OSVYK2;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 13-FEB-2005 (TrEMBLrel. 31, Last annotation update)
DE Collagen, type XII, alpha 1.
GN Name=COL12A1; ORFNames=RPL-23BD15.1-003;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Tromans A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Smith M.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Corby N.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL354664; CAH71309.1; -; Genomic DNA.
DR EMBL; AL080250; CAH19897.1; -; Genomic DNA.
DR EMBL; AL096771; CAH19907.1; -; Genomic DNA.
DR EMBL; AL080250; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL096771; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL354664; CAH19897.1; JOINED; Genomic DNA.
DR EMBL; AL080250; CAH19907.1; JOINED; Genomic DNA.

```

DR EMBL: AL354664; CA119897.1; JOINED; Genomic DNA.
 DR EMBL: AL096771; CA119897.1; JOINED; Genomic DNA.
 DR EMBL: ENSG00000111799.1; Homo sapiens.
 DR GO: GO:0005737; Cytoplasm; IEA.
 DR GO: GO:0005578; Extracellular matrix (sensu Metazoa); IEA.
 DR GO: GO:0005198; F-actin molecular activity; IEA.
 DR GO: GO:0007155; P-actin molecular activity; IEA.
 DR GO: GO:0006817; P-actin molecular activity; IEA.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR003961; FN III.
 DR InterPro: IPR003129; Laminin_G_TSP_N.
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF00041; Fn3; 18.
 DR Pfam: PF00092; VWA; 4.
 DR PRINTS: PRO0453; VWFADOMAIN.
 DR SMART: SM00210; TSPN; 1.
 DR SMART: SM00327; VWA; 4.
 DR PROSITE: PS00853; FN3; 18.
 DR PROSITE: PS0234; VWF; 4.
 DR Collagen; Extracellular matrix; Repeat; Structural protein.
 KW SEQUENCE 2884 AA; 315869 MW; 2D598F13656E5AD CRC64;

Alignment Scores:

Pred. No.: 1,92e-14 Length: 2884
 Score: 421.50 Matches: 120
 Percent Similarity: 46.5% Conservative: 67
 Best Local Similarity: 29.9% Mismatches: 186
 Query Match: 17.7% Indels: 29
 DB: 2 Gaps: 8

US-10-699-035a-5 (1-1254) x OSVYK2_HUMAN (1-2884)

QY 94 CGAGGGAGCTGATGTTCTCGTCGAGACGCTGCTCTCACTACAGTTCTCC 153
 Db LysAlaSpIleValIlePheLeuValIlePheGlySerTyrIleGlyIleAlaPheVal 457
 QY 154 CGGTTTCGAGAGTTTGGGGAGCTGTCCTCCACTGCGGAGCAGCGGAGCTG 213
 Db LysValAlaArgAlaPheLeuGluValLeuValLysSerPheGluLeuSerProAlaVal 477
 QY 214 CGTGCAGTCTGTGACGTCGTCGAGTCGCGGACGTCGAGTCGCTTCCCTTGGCAGCAC 273
 Db GlnIleSerLeuValGlnTyrSerArgSerProIleThrIlePheThrIleLysPhe 497
 QY 274 AGCTGGGTTGAGCTGCGGAGTCGTCGCTTCTCCAGCGCATGGTGACAC 333
 Db ThrLeuValIleGluPheIleGluAlaIleAsnThrPheProTyrArgIleGlySerThr 517
 QY 334 CACACTGGCTGCGGCTGCTATGCAAGAACAGCTGTTGCTGACATGACATGCGC 393
 Db AsnThrGlyLysAlaMetThrTyrValArgIleLysIlePheValProSerLysGlySer 537
 QY 394 CGGCGAGGGTGGCCAAAGTGTGTGTGGTGACAGATGGCGGCTCCAGCGACCTGTG 453
 Db ArgSerAsnValProLysValMetIleLeuIleThrAspIleLysSerSerAspAlaPhe 557
 QY 454 GCGCCCCCATGACAGACTCAAGAGACCTGGGCTGACCGTTCATTGTCAGACCGGC 513
 Db ArgAspProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValIleVal 577
 QY 514 CGAGGCAACTTCTGAGCTGTGACGCTGCGTCGCGGCTGCGGAGCGGAGCTGCGAC 573
 Db AspAlaValArgSerGluLeuGluAlaIleAlaSerProAlaGluThrIleValPhe 597
 QY 574 TTGTG---GACGTGATGACCTGACATCATTTGTCAGAGCTGAGGGCTTCATT--- 627
 Db ThrValGluPhePheAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCys 617
 QY 628 -----CTCGCG-----ATGCGCGCGGACGAG 648
 Db LeuArgIleGluGluGluLeuAlaIleLysLysLysLysLysLysLysLysLysLys 637

QY 649 CTCATGCCAGGAGTCACTGTCAGCGGCTTCCGCTGCGCTGCGCACCTCTGAC 708
 Db LeuSerPheSerGluValThrSerTyrGlyPheLysThrAsnThrSerProAlaGly 657
 QY 709 GCAGACTCGGCTACTATGCTGTCGAGCTGTGTGCCAGCGGCCAGCGGGGCTGAGA 768
 Db AsnValPheSerTyrIleIle-----ThrTyrLysGluAlaIleGly 671
 QY 769 CGCCAGCAGCTG-----CCAGGAAGCCAGCAGTGTGATGCTGAGCGGCTC 816
 Db AspAlaValAlaThrValValGluProAlaSerSerThrSerValValLeuSerSer 631
 QY 817 GACCGGACAGCAGCAGTCAAGCTGCGCTGAGCTGAGCTGCAAGCGGCTGAGG 876
 Db LysProGluThrLeuTyrLeuValAlaValThrAlaGluTyrGluLysPheSerIle 711
 QY 877 CCCAGATCTCGGGTGGCGCAGCGGCGCAGAGAGCGGCGCAGCGGCTGAGG 936
 Db ProLeuAlaGlyGluGluThrThrGluGluValLysGlyAlaProArgAsnLeuVal 731
 QY 937 TCCACAGCGCGCGCGCAGCTGCGCTGAGTGGGCGCCAGCGCTGAGCTGAGCGG 996
 Db ThrAspGluThrThrAspSerPheLysIleThrTyrThrGluAlaProIle-----Arg 749
 QY 997 GCGCTGCTGACACGCTGCACTGCGGCGCTGCGGCGGCGGCGGCGGCGGCTGAG 1056
 Db ValLeuArgTyrArgIleTyrArgProValAlaGlyGlyLysSerArgGluValThr 769
 QY 1057 GTGCCCGCGCGCGCAGCTGACACAGCTGCGAGGCTGAGCGCGCGGCTGAG 1116
 Db ThrProAlaGluArgArgArgThrLeuGluAsnLeuIleProAspThrTyrGlu 789
 QY 1117 GTGACCGTACCGCGCTTCCGCTGCGGCGGCGGAGAGCGGCTGCGGCGGCTGAG 1176
 Db ValSerValIleProGluTyrPheSerGlyProGlyThrProLeuThrGlyAlaIle 809
 QY 1177 AGCGCGAGCGGAGC 1236
 Db ThrGluGlu---ValArgLysAsnProArgAsnLeu---ArgValSerAspProThrThr 827
 QY 1237 GCCAGC 1242
 Db 828 SerThr 829

RESULT 12
 OSVYK1_HUMAN
 ID OSVYK1_HUMAN PRELIMINARY; PRT; 3063 AA.
 AC OSVYK1;
 DT 01-FEB-2005 (Tremblrel. 29, Created)
 DT 01-FEB-2005 (Tremblrel. 29, Last sequence update)
 DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
 DE Collagen, type XII, alpha 1
 GN Name=COL12A1; ORFNames=RPI-238D15.1-1001;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae;
 OC Homo.
 NC NCB1_TaxID=9606;
 RN RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Tromans A.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Smith M.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Corby N.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL354664; CA119897.1; -; Genomic DNA.
 DR EMBL: AL080250; CA119898.1; -; Genomic DNA.

DR EMBL, AL096771; CA119908.1; -, Genomic DNA.
 DR EMBL, AL088250; CAH71310.1; JOINED; Genomic DNA.
 DR EMBL, AL096771; CAH71310.1; JOINED; Genomic DNA.
 DR EMBL, AL354664; CA119898.1; JOINED; Genomic DNA.
 DR EMBL, AL088250; CA119908.1; JOINED; Genomic DNA.
 DR EMBL, AL354664; CA119908.1; JOINED; Genomic DNA.
 DR EMBL, AL096771; CA119898.1; JOINED; Genomic DNA.
 DR EMBL, ENSG0000111799; Homo sapiens.
 DR GO, GO:0005737; Cytoplasm; IEA.
 DR GO, GO:0005578; Extracellular matrix (sensu Metazoa); IEA.
 DR GO, GO:0005198; P:structural molecule activity; IEA.
 DR GO, GO:0007155; P:cell adhesion; IEA.
 DR GO, GO:0006817; P:phosphate transport; IEA.
 DR InterPro, IPR008160; Collagen.
 DR InterPro, IPR003961; FN_III.
 DR InterPro, IPR002035; VWF A.
 DR Pfam, PF01391; Collagen; 4.
 DR Pfam, PF00041; fn3; 18.
 DR Pfam, PF00092; VWA; 4.
 DR PRINTS, PR00453; VWFADOMAIN.
 DR SMART, SM00060; FN3; 18.
 DR SMART, SM00210; TSPN; 1.
 DR SMART, SM00327; VWA; 4.
 DR PROSITE, PS50853; FN3; 18.
 DR PROSITE, PS50334; VWF A; 4.
 DR KX, Collagen, Extracellular matrix; Repeat; Structural protein.
 DR KX, SEQUENCE 3063 AA; 333147 MW; EA38CAFECB8393D2 CRC64;

Alignment Scores:

Pred. No.: 1,92e-14 Length: 3063
 Score: 421.50 Matches: 120
 Percent Similarity: 46.5% Conservative: 67
 Best Local Similarity: 29.9% Mismatches: 186
 Query Match: 17.7% Indels: 29
 DB: 2 Gaps: 8

US-10-699-035a-5 (1-1254) x Q5VYK1 HUMAN (1-3063)

QY 94 CGAGGGAGCTGATGTTCTGCTGACGACTGACCGAGCTGCTCTACAGATTCTCC 153
 DB 438 LysAlaAspIleValPheLeuValAspGlySerTyrSerIleGlyIleAlaAsnPheVal 457
 QY 154 CGGGTTCGGAGATTGTGGGGAGAGCTGTGGCTCACTCCCTGGAGCCGGGGCCCTG 213
 DB 458 LysValAlaArgAlaPheLeuValIleValLysSerPheGluIleSerProAsnArgVal 477
 QY 214 CGTGCAGCTGCTGACGCTGGGAGCTGGCCATACCGGATCCCTTCCGGCCAGCAC 273
 DB 478 GlnIleSerLeuValGlnTyrSerArgAspProH1stHgluPheThrLeuLysSph 497
 QY 274 AGCTCGGGTGAAGCTGCGCCAGATGCGGTGCTTCTGCCAGCGATGGGTGACACC 333
 DB 498 ThrLysValGluAspIleIleGluAlaIleAsnThrPheProTyrArgIlyGlySerThr 517
 QY 334 CACACTGCGCTGGCGCTGTCTATGCAAGAAACAGCTGTTGCTGAACCATCAGGTGCC 393
 DB 518 AsnThrGlyLysAlaIleMetThrTyrValArgGlyLysIlePheValProSerIysGlySer 537
 QY 394 CGGCGACGGGGTCCCAAGTGTGCTGTGGTGAAGACAGAGGGGCTCCGACCCCTG 453
 DB 538 ArgSerAsnValProLysValMetIleLeuIleThrAspGlyLysSerSerAspAlaPhe 557
 QY 454 GGGCCGCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCGTTCATTGTTCAGACCGGC 513
 DB 558 ArgAspProAlaIleIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValLys 577
 QY 514 CGAGGCAACTTCTCTGAGCTGTCAAGCGCTGCTTCCAGCCCTGCGCAGAAAGACCTGCAC 573
 DB 578 AspAlaValArgSerGluLeuGluAlaIleAlaSerProProAlaGluThrIstValPhe 597
 QY 574 TTTGTC---GAGCTGATGATCACTGCACATCATTTGTCGAAGACTGAGGGGCTCCATT--- 627

DB 598 ThrValGluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCys 617
 QY 628 -----CTCGCG-----ATCGCGCGCAGCAG 648
 DB 618 LeuArgIleGluGlnGluIleAlaIleIleLysLysValTyrValProLysAsp 637
 QY 649 CTCATGCGCAGGAAATACATCTCCAGCTTCGCGCTGGCTGGCCACCCCTGCTGACC 708
 DB 638 LeuSerPheSerGluValThrSerTyrGlyPheLysThrAsnTrpSerProAlaGlyLeu 657
 QY 709 GCAGACTGGGCTACTATGCTGTGAGCTGGGCCAGCGCCAGCGGGGCTGCACAGA 768
 DB 658 AsnValPheSerTyrHisIle-----ThrTyrLysGluAlaIleGly 671
 QY 769 CGCCAGCAGCTG-----CCAGGAACCGCACGAGCTGATCTGGCGGCGCTC 816
 DB 672 AspArgGluValThrValValGluProAlaSerSerThrSerValValLeuSerLeu 691
 QY 817 GACCCGACAGGACTACAGCTGGCGCTGAGCTGAGTCCAGCTGCGCTCTCTGAGG 876
 DB 692 LysProGluThrLeuTyrLeuValAsnValThrAlaGluTyrGluAspGlyPheSerIle 711
 QY 877 CCCAGATCTGCGGGTGGCGCAGCGCGCAGAGAGCGCGGCGCAGCGATCGTCAN 936
 DB 712 ProLeuAlaGlyGluGluThrThrGluGluValLysGlyAlaProArgAsnLeuVal 731
 QY 937 TCCACGCGCGCGCGCAGCTCCGCTGAGATTGGGCCCGCAGCGCTGAGCTCAGCGCG 996
 DB 732 ThrAspGluThrThrAspSerPheLysIleThrTrpThrGlnAlaProGly-----Arg 749
 QY 997 GCGCTCGGCTACACGTGAGTTCGGCGCTGCGGGGCGGGGAGCGCAGCGGTGAG 1056
 DB 750 ValLeuArgTyrArgIleIleTyrArgProValAlaGlyGlyLysSerArgGluValThr 769
 QY 1057 GTGCCGCGGGCGGCACTGACACCGCTGCGGAGGCTGGCGCGGCGGCGCAGCTG 1116
 DB 770 ThrProProAsnGlnArgArgThrLeuGluAsnLeuIleProAspThrLysTyrGlu 789
 QY 1117 GTGACCTGACCGCGCTTCCGCTCGGCGCGCAGAGCGCGCTGTCGCCAAGGCTGC 1176
 DB 790 ValSerValIleProGluTyrPheSerGlyProGlyThrProLeuThrGlyAsnAlaIle 809
 QY 1177 AGCGCCGAGCG 1236
 DB 810 ThrGluGlu---ValArgGlyAsnProArgAspLeu---ArgValSerAspProThrThr 827
 QY 1237 GCCAGC 1242
 DB 828 SerThr 829

RESULT 13
 COCAL HUMAN STANDARD; PRT; 3063 AA.
 ID COCAL HUMAN 099715; 099716;
 AC 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor.
 GN Name=COL12A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT), AND PARTIAL PROTEIN
 RP MEDLINE=97288521; PubMed=9143499; DOI=10.1006/geno.1997.4638;
 RA Gersheke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,
 RA Hudson D.L., Champilaud M.-F., Olsen B.R., Burgess R.E.;
 RT "Complete primary structure of two splice variants of collagen XII,
 RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(XI)
 RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human

RT chromosome 6q12-q13.
 CC Genomics 41:236-242 (1997).
 CC - FUNCTION: Type XII collagen interacts with type I collagen-
 CC containing fibrils, the COL1 domain could be associated with the
 CC surface of the fibrils, and the COL2 and NC3 domains may be
 CC localized in the pericellular matrix (By similarity).
 CC - SUBUNIT: Trimer of identical chains each containing 190 kDa of
 CC nontriple-helical sequences.
 CC - ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Comment=The final tissue form of collagen XII may contain
 CC homotrimers of either isoform Long or isoform Short or any
 CC combination of isoform Long and isoform Short;
 CC Name=Long;
 CC IsoId=Q09715-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=Q09715-2; Sequence=VSP_001149;
 CC - TISSUE SPECIFICITY: Found in collagen I-containing tissues: both
 CC isoform Short and isoform Long appear in amnion, chorion, skeletal
 CC muscle, small intestine, and in cell culture of dermal
 CC fibroblasts, keratinocytes and endothelial cells. Only the short
 CC isoform is found in lung, placenta, kidney and a squamous cell
 CC carcinoma cell line.
 CC - PTM: The triple-helical tail is stabilized by disulfide bonds at
 CC each end (By similarity).
 CC - PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains (By
 CC similarity).
 CC - PTM: O-glycosylation of isoform Long; glycosaminoglycan of
 CC chondroitin-sulfate type (By similarity).
 CC - SIMILARITY: Belongs to the fibril-associated collagens with
 CC interrupted helices (FACIT) family.
 CC - SIMILARITY: Contains 18 fibronectin type-III domains.
 CC - SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC - SIMILARITY: Contains 4 VWFA domains.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL, U73778; AAC51244.1; -; mRNA.
 CC EMBL, U73779; AAD40483.1; -; mRNA.
 CC HSSP: P18614; IMHP.
 CC DR HSSP: ENSG0000011799; Homo sapiens.
 CC DR HGN: HGNC:2188; COL12A1.
 CC DR MIM: 120320; -.
 CC DR GO: GO:0005595; C:collagen type XII; TAS.
 CC DR GO: GO:0001501; P:skeletal development; TAS.
 CC DR InterPro: IPR008160; Collagen.
 CC DR InterPro: IPR003961; FN_III.
 CC DR InterPro: IPR002035; VF_A.
 CC DR Pfam: PF01391; Collagen; 4.
 CC DR Pfam: PF00092; VWA; 4.
 CC DR PRINTS: PR00453; VWFADOMAIN.
 CC DR PROSITE: PS00853; FN3; 18.
 CC DR PROSITE: PS00234; VWA; 4.
 CC DR Alternative splicing; Cell adhesion; Collagen;
 CC Direct protein sequencing; Extracellular matrix; Glycoprotein;
 CC Hydroxylation; Repeat; Signal; Structural protein.
 CC FT SIGNL 1 24 Potential.
 CC FT CHAIN 25 3063 Collagen alpha 1(XII) chain.
 CC FT DOMAIN 25 112 Fibronectin type-III 1.
 CC FT DOMAIN 140 316 VWFA 1.
 CC FT DOMAIN 333 422 Fibronectin type-III 2.
 CC FT DOMAIN 440 616 VWFA 2.
 CC FT DOMAIN 631 719 Fibronectin type-III 3.
 CC FT DOMAIN 722 810 Fibronectin type-III 4.
 CC FT DOMAIN 813 901 Fibronectin type-III 5.
 CC FT DOMAIN 904 993 Fibronectin type-III 6.
 CC FT DOMAIN 995 1083 Fibronectin type-III 7.

FT	DOMAIN	1086	1175	Fibronectin type-III 8.
FT	DOMAIN	1199	1371	VWFA 3.
FT	DOMAIN	1384	1472	Fibronectin type-III 9.
FT	DOMAIN	1474	1563	Fibronectin type-III 10.
FT	DOMAIN	1565	1652	Fibronectin type-III 11.
FT	DOMAIN	1654	1743	Fibronectin type-III 12.
FT	DOMAIN	1752	1841	Fibronectin type-III 13.
FT	DOMAIN	1843	1931	Fibronectin type-III 14.
FT	DOMAIN	1933	2022	Fibronectin type-III 15.
FT	DOMAIN	2024	2113	Fibronectin type-III 16.
FT	DOMAIN	2115	2202	Fibronectin type-III 17.
FT	DOMAIN	2206	2280	Fibronectin type-III 18.
FT	DOMAIN	2323	2496	VWFA 4.
FT	DOMAIN	2520	2712	TSP N-terminal.
FT	DOMAIN	2451	2746	Nonhelical region (NC3).
FT	REGION	2747	2898	Triple-helical region (COL2) with 1
FT	REGION	2899	2941	imperfector.
FT	REGION	2942	3044	Nonhelical region (NC2).
FT	REGION	3045	3063	Triple-helical region (COL1) with 2
FT	REGION	3064	3071	imperfector.
FT	MOTIF	862	864	Cell attachment site (Potential).
FT	MOTIF	2779	2781	Cell attachment site (Potential).
FT	MOTIF	2895	2897	Cell attachment site (Potential).
FT	MOD_RES	2944	2944	Hydroxyproline (By similarity).
FT	MOD_RES	2947	2947	Hydroxyproline (By similarity).
FT	MOD_RES	2950	2950	Hydroxyproline (By similarity).
FT	MOD_RES	2959	2959	Hydroxyproline (By similarity).
FT	MOD_RES	2965	2965	Hydroxyproline (By similarity).
FT	MOD_RES	2968	2968	Hydroxyproline (By similarity).
FT	MOD_RES	2971	2971	Hydroxyproline (By similarity).
FT	MOD_RES	2983	2983	Hydroxyproline (By similarity).
FT	MOD_RES	3000	3000	Hydroxyproline (By similarity).
FT	MOD_RES	3003	3003	Hydroxyproline (By similarity).
FT	MOD_RES	3014	3014	Hydroxyproline (By similarity).
FT	MOD_RES	3023	3023	Hydroxyproline (By similarity).
FT	MOD_RES	3026	3026	Hydroxyproline (By similarity).
FT	MOD_RES	3029	3029	Hydroxyproline (By similarity).
FT	CARBOHYD	700	700	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	798	798	O-linked (Xyl. . .) (chondroitin sulfate)
FT	CARBOHYD	889	889	(Potential).
FT	CARBOHYD	981	981	O-linked (Xyl. . .) (chondroitin sulfate)
FT	CARBOHYD	1763	1763	(Potential).
FT	CARBOHYD	2206	2206	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	2528	2528	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	2679	2679	N-linked (GlcNAc. . .) (Potential).
FT	VARSPLIC	25	1188	Missing (in isoform Short).
FT	SEQUENCE	3063 AA;	333194 MW;	75FEA78FA8E8253 CRC64;

Alignment Scores:
 Pred. No.: 2,6e-14
 Score: 419.00
 Percent Similarity: 39.9%
 Best Local Similarity: 27.5%
 Query Match: 17.6%
 DB: 1
 Gaps: 9

US-10-699-035A-5 (1-1254) x COCAL_HUMAN (1-3063)
 94 CGAGGAGCTGATGTTCTGTCGACAGCCAGGCTCTCATTACGAGTTCTCC 153
 Db 438 LySAlaApIleValPheLeuValApGlySerTyrSerIleValAlaenPheVal 457
 Oy 154 CGAGTTCGAGAGTTGTGGGAGCTGTGCTCCATCCCTCGGAGCCGCGCTG 213
 Db 458 LySValArgAlaPheLeuValIleValValSerPheGluIleSerProAsnArgVal 477
 Oy 214 CGTGCAGTCTGGTGCACGTGGAGCCATACCGAGTTCCTTCGCGCCAGCAC 273


```

Db      478 GlnIleSerLeuValGlnThrSerArgAspProHisThrGluPheThrLeuLysPhe 497
QY      274 AGCTGGGTGAGAGCTGCCAGGATGGGTGCTTCTGCCAGCGACGTGGACAC 333
Db      498 ThrLysValGluAspLeuIleGluAlaIleAsnThrPheProTyrArgIlyGlySerThr 517
QY      334 CACACTGGCTGCGCTGTGCTATGCGCAAGAACAGCTTTGCTGACATCAGCTGCC 393
Db      518 AsnThrGlyLysAlaMetThrTyrValArgGlyLysIlePheValProSerLysGlySer 537
QY      394 CGGCGCAGGGGGTCCCAAGTGGGTGGTGGAGATGGGGGCTCCGACCGACCTGTG 453
Db      538 ArgSerAsnValProLysValMetCileuIleHisAspGlyLysSerSerAspAlaPhe 557
QY      454 GGCCTCCCATGACAGAGCTCAAGACCTGGGCGTCACTGCTTCAATTTGACACCGGC 513
Db      558 ArgAspProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyLys 577
QY      514 CGAGGCACTTCTGTGAGCTGTCAAGCTGCTGCTGACGCTGCGGACGACACCTGCAC 573
Db      578 AspAlaValArgSerGluLeuGluAlaIleAlaSerProAlaGluThrHisValPhe 597
QY      574 TTGTGTG---GACGTGGATGACCTGCACATCATTTGTCAGAGCTGAGGGGCTCCATT--- 627
Db      598 ThrValGluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCys 617
QY      628 -----CTCGCG-----ATGCGCGCGCAGCAG 648
Db      618 LeuArgIleGluGlnLeuAlaIleLysLysValAlaIleTyrValProProLysAsp 637
QY      649 CTCATGCCACGAGATCACGTCACGCGCTTCCGCTGCGCTGCGACCCCTG----- 702
Db      638 LeuSerPheSerGluValThrSerTyrGlyPheLysThrAsnTrpSerProAlaGlyGlu 657
QY      703 -----CTACCGCAGACCTGGGCTACTGATG 729
Db      656 AsnValPheSerTyrHisIleThrTyrLysGluAlaAlaGlyAspAspGluValThrVal 677
QY      730 CTGAG-----CTGCTGCCAGCGCGCCAGCG----- 756
Db      678 ValGluProAlaSerSerThrSerValValLeuSerSerLeuLysProGluThrLeuTyr 697
QY      756 ----- 756
Db      698 LeuValAsnValThrAlaGlyLysArgLysPheSerIleProLeuAlaGlyGluGln 717
QY      757 -----GGGCTGCACAGCGCGCACAGCTGCACAGGACGACGACGAC 798
Db      718 ThrThrGluGluValLysGlyAlaProArgAsnLeuLysValThrAspGluThrThrAsp 737
QY      799 -----TGATCTGGGCC----- 810
Db      738 SerPheLysIleThrTrpThrGlnAlaProGlyArgValLeuArgCysArgIleIleTyr 757
QY      810 ----- 810
Db      758 ArgProValAlaGlyLysLysSerArgLysValThrThrProProAsnGlnArgArgArg 777
QY      811 -----GGCTGCACCGCGCAGGACGACTACGAGTGGCGCTAGTGGCTGAGTCACAC 861
Db      778 ThrLeuGluAsnLeuIleProAspThrLysTyrGluValSerValIleProGluTyrPhe 797
QY      862 GTGCGCTCTCTGAGCGCCAGATCTCGCGGTGCGACGCGCGCAGAGAGCGCGGCGCA 921
Db      798 SerGlyProGlyThrProLeuThrGlyAsnAlaAlaThrGluGluValArgLysAsnPro 817
QY      922 GAGCGCATGCTATCTCCACGCGCGCGCGCAGCGCTCCGCTGAGTGGCGCGCGCAGCG 981
Db      818 ArgAspLeuArgValSerAspProThrThrSerThrMetLysLeuSerTrpSerGlyAla 837
QY      982 CTGGGCTCAGCGCGCGCTCGGCTACACGCTGACATGCGCGCGCTGCGCGCGCGGAG 1041
Db      838 ProGlyLysValLysGln-----TyrLeuValThrTyrThrProValAlaGlyGlyGlu 855

```

```

QY      1042 GCGCAGCGGTGAGAGTCCCGCGCGCGCAACTGCACCTGCAGGCGCTTGGCGCG 1101
Db      856 ThrThrGluValThrValArgGlyAspThrThrAsnThrValLeuGlnGlyLeuLysGlu 875
QY      1102 GGCACCGGCTTACTGTGTATACCGTGAACCGCGCTTCCGCTGCGCGCGGACGCGCGCTG 1161
Db      876 GlyThrGlnTyrAlaLeuSerValThrAlaLeuTyrAlaSerGlyAlaGlyAspAlaLeu 895
QY      1162 TCCGCGCAAGCGCTGCACG 1179
Db      896 PheGlyGluGlyThrThr 901

RESULT 14
Q4SD22 TETNG
ID Q4SD22 TETNG PRELIMINARY; PRT; 1723 AA.
AC Q4SD22
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DE Chromosome 14 SCAF14645, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00020231001;
OS Tetradodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Acanthopterygii; Percormorpha; Tetradodontiformes;
OC Tetradontidae; Tetradontidae; Tetradodon.
OX NCBI_Taxid=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicoud S., Jaffe D., Fisher S., Lutfalla G., Dessat C., Segurens B.,
RA Dastiva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Catcolico L., Poulat J., De Bernardis V.,
RA Bieumont C., Skallil Z., Catcolico L., Poulat J., De Bernardis V.,
RA Cruad C., Duprat S., Broctier P., Coutanceau J.P., Guzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Querier F., Saurin W., Scarpelli C.,
RA Winkler P., Lander E.S., Weissbach J., Roest Crolious H.,
RA "Genome duplication in the teleost fish Tetradodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope, Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL, CAAB01014645; CAG01460.1; -, Genomic_DNA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_subd.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00041; fn3; 11.
DR Pfam; PF00092; VWA; 3.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 11.
DR SMART; SM00327; VWA; 3.
DR PROSITE; PS00853; FN3; 12.
DR PROSITE; PS50234; VWFA; 3.
KW Collagen; Extracellular matrix; Structural protein.
FT NON_TER 1
FT NON_TER 1723
SQ SEQUENCE 1723 AA; 187496 MW; 7EB78DA80D1E6B40 CRC64;

Alignment Scores:
Pred. No.: 6,47e-14 Length: 1723
Score: 411.50 Matches: 116

```


Percent Similarity: 44.6%
 Best Local Similarity: 30.3%
 Query Match: 17.3%
 DB: 2
 Gaps: 5

US-10-699-035A-5 (1-1254) x QASD22_TETNG (1-1723)

QY 94 CGAGGGACCTGATGTTCTGTGACAGCTGACGCGCTCTGACAGAGTTCTCC 153
 DB 436 GlnAlaaspValValLeuValValaspGlySerTyrSerIleGlyLeuAlaenPheAla 455
 QY 154 CGGGTTCGGAGCTTTGGGGAGCTGGTGTCCACTGCCCCGGGACCGGGGCTTG 213
 DB 456 LysValAlaAlaPheLeuGluValLeuValAlaSerThrPheAspIleGlyProAlaVal 475
 QY 214 CGTCCAGCTGTGATGACGCGGACGCGGATACAGGATCCCTTCCGCGACAC 273
 DB 476 GlnIleSerLeuValGlnTyrSerAlaPheProIleThrGluPheTyrLeuAspSerHis 495
 QY 274 AGCTCGGATGAGCTGCCAGATGCGGCGCTTGTGCCAGCGCATGGGTGACACC 333
 DB 496 HisAsnLeuGlnAlaValValThrAlaLeuArgThrPheProTyrArgGlySerThr 515
 QY 334 CACACTGGCTGCGCTGCTCTTGCACAGAAACAGCTTTGCTGAAGCATCAGTCC 393
 DB 516 AsnThrGlyAlaAlaMetThrTyrValAlaGluThrValPheGlnAlaSerArgGlyAla 535
 QY 394 CGGCGCGGGGTGCGCAAGTGTGCTGTGTCAGACGCGGCTCCAGCGACCTGTG 453
 DB 536 ArgAlaHisValAlaProAlaValThrIleLeuIleThrAspGlyLysSerSerAlaPhe 555
 QY 454 GCGCCCCCATGACGAGATCAAGACACTGCGCGCTGACGCTTCTTGTGACACCGGC 513
 DB 556 GlnGluProAlaAlaAsnLeuArgAsnSerAlaValGlnIlePheAlaValGlyValys 575
 QY 514 CGAGGACACTTCTGAGCTGTGATGCGGCTGACGCGCTCCGCGAGAACACTGAC 573
 DB 576 AspAlaValAlaArgSerGluLeuGlnAlaIleAlaAsnAlaProAlaGluThrHisValTyr 595
 QY 574 TTGTG---GACGTGATGACCTGACATCATGCTCCAGAGTGGGGCTCATCTCC 630
 DB 596 ThrValGluAspPheAspAlaPheGlnArgIleSerThrGluLeuThrGlnSerIleCys 615
 QY 631 -----GCGATGCGCGCGACGAC 648
 DB 616 LeuArgIleGluGlnGluLeuGlnIleIleAsnGlnArgAlaLeuValGlnProArgAsp 635
 QY 649 CTCATGACGAGATCACTGACGCGGCTTCCGCTGCGCTGCGCACTGCTGACC 708
 DB 636 LeuTyrPheSerAspValGlyProArgSerPheArgAlaSerTrpGluIleAsnAlaAsn 655
 QY 709 GCGAGCTCGGGCTACATGCTGCGAGCTGTGTCCAGCGCGCGGGGCTGCAAA 768
 DB 656 AsnValGluSerTyrLeuValGlnPheArgProThrGluGlyValLysPheSerHisTyrVal 675
 QY 769 CGCCAGACACTGCGCAGGACCGCACGAGCTGATGCGCGCTGCGACCGGACG 828
 DB 676 SerMetSerValProLysAspValLeuThrAlaLeuLeuProHisLeuThrProLeuThr 695
 QY 829 GACTACGACGTGCGCTAGTCTGATGCCACGTCAGCGCTCTGAGGCCCGCATCTG 888
 DB 696 ArgTyrGluValSerValSerAlaGlnTyrAlaTyrGlyThrSerLeuProValThrGly 715
 QY 889 CGGTCGCGACCGGCGGACGAGCGCGCGACGCGATCTCATCTCCACGCGCG 948
 DB 716 TyrGlyThrThrAlaGluGlnArgGlySerValGlnAsnLeuValThrGluGluSer 735
 QY 949 CGCGGAGCTCGCGCTGATGAGGCGCGACGCTGCGCTGACCGCGCGCTCGGCTAC 1008
 DB 736 ProGlnSerPheArgValSerTrpAlaAlaProGlyValAlaAlaArg-----Tyr 753
 QY 1009 CACGTCACTTGGCGCGCTGCGGGCGGGAGCGCGCGGTGAGGTGCGCGCGCG 1068

DB 754 ArgLeuThrTyrGlnProAlaGlyValGlyAlaGlnLeuGlnAlaPheThrAlaGly 773
 QY 1069 CGCACTGACAC---ACGCTGACAGGCGCTGAGCGCGGACCGGCTACTGTGACCGTG 1125
 DB 774 ProGluLeuThrMetValLeuGlnIleAsnLeuGlnProArgThrThrTyrArgValThrVal 793
 QY 1126 ACCGCGCTTCCGCTCGGGCGCGAGACGCGCTGTCCGCAAGGCTGACGCGCGAC 1185
 DB 794 ThrProGluTyrGlyGly-----ProGly 802
 QY 1186 GCGCGCGCG 1194
 DB 803 GlyProGln 805

RESULT 15
 COCAL MOUSE
 ID COCAL MOUSE STANDARD; PRT; 3119 AA.
 AC 060847; P70322;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor.
 GN Name=Col12a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND XIIA-2).
 RP STRAIN=C57BL/6J, and Swiss Webster; TISSUE=Skin;
 RX MEDLINE=96170761; PubMed=8601036;
 RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
 RT "Primary structure of the long and short splice variants of mouse collagen XII and their tissue-specific expression during embryonic development.";
 RL Dev. Dyn. 204:432-445(1995).
 RN [2]
 RP PARTIAL NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XIIA-2 AND XIIA-1).
 RP STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
 RX MEDLINE=95348349; PubMed=10419532; DOI=10.1074/jbc.274.31.22053;
 RA Kana A.M., Reichenberger E., Baur S.T., Karimbox N.Y., Taylor R.W., Olsen B.R., Nishimura I.;
 RT "Structural variation of type XII collagen at its carboxyl-terminal NCI domain generated by tissue-specific alternative splicing.";
 RL J. Biol. Chem. 274:22053-22059(1999).
 CC -I- FUNCTION: Type XII collagen interacts with type I collagen-containing fibrils, the COL1 domain could be associated with the surface of the fibrils, and the COL2 and NC3 domains may be localized in the pericellular matrix (by similarity).
 CC -I- SUBUNIT: Trimer of identical chains each containing 190 kDa of nontriple-helical sequences (by similarity).
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=The final tissue form of collagen XII may contain homooligomers or any combination of the various isoforms;
 CC Name=XIIA-1;
 CC IsoId=Q60847-1; Sequence=Displayed;
 CC Name=XIIA-2; Synonyms=ER#X;
 CC IsoId=Q60847-2; Sequence=VSP_001151, VSP_001152;
 CC Name=XIIB-1;
 CC IsoId=Q60847-3; Sequence=VSP_001150;
 CC Name=XIIB-2;
 CC IsoId=Q60847-4; Sequence=VSP_001150, VSP_001151, VSP_001152;
 CC -I- TISSUE SPECIFICITY: Highest expression in tendons, perichondrium, skin, cornea, sclera, blood vessels, and peritoneum.
 CC -I- DEVELOPMENTAL STAGE: The long NC3 XIIA isoforms are predominant at early stages (E07 and 11); at later stages of development (E015 and 17) the short NC3 XIIB forms become the major forms. As the short NC3 forms become the major product, the long splice variant continues to be expressed in several tissues, even after birth.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:33:59 ; Search time 142.074 Seconds

(without alignments)
893.866 Million cell updates/sec

Title: US-10-699-035a-2

Perfect score: 913

Sequence: 1 RGDLMFLDSSASVSHYFES.....FVDVDDHLITVQLRGSLID 180

Scoring table: BLOSUM62

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	913	100.0	445	2	Q6PCB0 HUMAN
2	772	84.6	415	2	Q8R2Z5 MOUSE
3	768	84.1	415	2	Q923K3 MOUSE
4	768	84.1	415	2	Q8C007 MOUSE
5	761	83.4	415	2	Q642A6 RAT
6	397	43.5	505	2	Q4SCD1 TETNG
7	273.5	30.0	1723	2	Q4SD22 TETNG
8	273.5	30.0	2225	2	Q4SXK3 TETNG
9	268.5	29.4	1259	2	Q4RP12 TETNG
10	265.5	29.1	1888	1	COEAL1 CHICK
11	265.5	29.1	3124	1	COEAL1 CHICK
12	265	29.0	624	1	MATN4 MOUSE
13	265	29.0	1557	2	Q4SH63 TETNG
14	263.5	28.9	1884	2	Q5VYK2 HUMAN
15	263.5	28.9	3063	1	COEAL1 HUMAN
16	263.5	28.9	3063	2	Q5VYK1 HUMAN
17	259	28.4	517	1	MATN4 HUMAN
18	257.5	28.2	517	2	Q4B853 HUMAN
19	256.5	28.1	1796	1	COEAL1 HUMAN
20	254.5	27.9	637	2	Q8IVX1 HUMAN
21	254.5	27.9	1284	2	Q4VX05 HUMAN
22	254.5	27.9	1284	2	Q4VX05 HUMAN
23	254.5	27.9	1284	2	Q4VX04 HUMAN
24	254.5	27.9	1284	2	Q4VX04 HUMAN
25	251.5	27.5	451	1	KI510 HUMAN
26	251.5	27.5	451	2	Q8N2G3 HUMAN
27	251	27.5	688	2	Q4T0K3 TETNG
28	250.5	27.4	839	2	Q5R9N1 PONY
29	250	27.4	839	2	Q6P3N7 XENTR
30	248.5	27.2	1117	2	Q4RXN8 TETNG
31	248.5	27.2	1117	2	Q4RXN8 TETNG

32	248	27.2	3119	1	COEAL1 MOUSE
33	247.5	27.1	493	1	MATN1 CHICK
34	246.5	27.0	644	2	Q5NJI1 BRARE
35	246.5	27.0	821	2	Q6PYX2 BRARE
36	246.5	27.0	944	2	Q5NJI5 BRARE
37	245.5	26.9	534	2	Q4G0W3 HUMAN
38	243.5	26.7	685	2	Q5NJI2 BRARE
39	242.5	26.6	937	2	Q569V0 MOUSE
40	242.5	26.6	956	1	MATN2 MOUSE
41	242.5	26.6	956	2	Q8R542 MOUSE
42	242	26.5	490	2	Q7SYT5 XENTR
43	241.5	26.5	647	2	Q4S2X7 TETNG
44	241.5	26.5	1450	2	Q4RP14 TETNG
45	240.5	26.3	500	2	Q80VNS MOUSE

ALIGNMENTS

RESULT 1	
Q6PCB0 HUMAN	
ID Q6PCB0 HUMAN PRELIMINARY; PRT; 445 AA.	
AC Q6PCB0:	
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)	
DT 13-SEP-2005 (TREMblrel. 31, Last annotation update)	
DE Von Willebrand factor A domain-related protein, isoform 1 (WARP).	
GN Name=WARP; ORFNames=RP4-758J18.11-001;	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;	
OC Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RC NUCLEOTIDE SEQUENCE.	
RP TISSUE=Placenta;	
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;	
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,	
RA Klausner R.D., Collins F.S., Wagner L., Schein C.M., Schuler G.D.,	
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,	
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,	
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,	
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaeetz T.E.,	
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,	
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,	
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,	
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,	
RA Villalón D.K., Muzny D.M., Sodergren E.V., Lu X., Gibbs R.A.,	
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,	
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,	
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,	
RA Butterfield V.S.N., Krzywinski M.I., Skalske U., Smalls D.E.,	
RA Scherf A., Schein J.E., Jones S.J.W., Marra M.A.,	
RT "Generation and initial analysis of more than 15,000 full-length human	
and mouse cDNA sequences."	
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).	
RN [2]	
RC NUCLEOTIDE SEQUENCE.	
RP TISSUE=Placenta;	
RA Strausberg R.L.	
RN [3]	
RC NUCLEOTIDE SEQUENCE.	
RP Hall R.;	
RA Submitted (May-2005) to the EMBL/Genbank/DBJ databases.	
DR EMBL; BC059409; AAHS9409.1; -; mRNA.	
DR EMBL; AL391244; CA12657.1; -; genomic DNA.	
DR Ensembl; ENSG00000179403; Homo sapiens.	
DR InterPro; IPR003961; FN III.	
DR InterPro; IPR002035; VWF_A.	
DR Pfam; PF00041; fn3; 2.	
DR Pfam; PF00092; VWA; 1.	

DR PRINTS; PR00453; VMPADOMAIN.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS00234; VMA; 1.
SO SEQUENCE 445 AA, 46804 MW, D9FBD83C0A4DEBDF CRC64;

Query Match 100.0%; Score 913; DB 2; Length 445;
Best Local Similarity 100.0%; Pred. No. 2,7e-72;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RGDMLFLDSSASVSHYERSRREFFVQGLVAPLPGTGLRSLVHNGSRPTTFRPQCH 60
DB 32 RGDMLFLDSSASVSHYERSRREFFVQGLVAPLPGTGLRSLVHNGSRPTTFRPQCH 91
OY 61 SSGEAAQDAVRAAQRMGDTHTGLALVYAKEQLFAEASGAPGPKVLVMTDGGSSDPV 120
DB 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKEQLFAEASGAPGPKVLVMTDGGSSDPV 151
OY 121 GPMQELKDLGVTFTVSTGRGNFLELSAASAPAEKHLFPVDVDDLHIIVQELRGSLID 180
DB 152 GPMQELKDLGVTFTVSTGRGNFLELSAASAPAEKHLFPVDVDDLHIIVQELRGSLID 211

RESULT 2
OBR225 MOUSE
ID OBR225 MOUSE PRELIMINARY; PRT; 415 AA.
AC OBR225;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 13-SEP-2005 (TREMblrel. 31, Last sequence update)
DE Von Willebrand factor A domain-related protein (Mus musculus 11 days
DE pregnant adult female ovary and uterus cDNA, RIKEN full-length
DE enriched library, clone:5031410123 product:VON WILLEBRAND FACTOR A-
DE RELATED PROTEIN homolog).
CN Name=Vval;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II, and FVB/N;
RC TISSUE=Colon, and
RC Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.M., Garcia A.M., Gay L.J., Hilyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marz M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX NIH GGC Project;
RG Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.

RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.",
RT Meth. Enzymol. 303:19-44(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi U., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski B., Yoshida K., Hasegawa Y., Kawai H., Kohlschütter S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.",
RT Nature 409:685-690(2001).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RT Nature 420:563-573(2002).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata K., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.",
RT Genome Res. 10:1617-1630(2000).
RN [7]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Katsunari T., Tashiro H., Itoh M.,
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaishiji K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohtsuka E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka K., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.",
RT Genome Res. 10:1757-1771(2000).
RN [8]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Nunazaki R., Ono M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,

RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RN Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 [9]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RA Director MGC Project;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC026919; AAH26919.1; -; mRNA.
 DR EMBL; AK077240; BAC36703.1; -; mRNA.
 DR EMBL; BC036166; AAH36166.1; -; mRNA.
 DR HSSP; P18614; IMHP.
 DR Ensembl; ENSMUSG0000042116; Mus musculus.
 DR MGI; MGI:2179729; Vwal.
 DR GO; GO:0005615; Cytoplasmic space; TAS.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00041; Fn3; 2.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00853; FN3; 2.
 DR PROSITE; PS0234; VWF_A; 1.
 DR PROSITE; PS0234; VWF_A; 1.
 SQ SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;

Query Match 84.6%; Score 772; DB 2; Length 415;
 Best local Similarity 83.9%; Pred. No. 7.5e-60;
 Matches 151; Conservative 13; Mismatches 16; Indels 0; Gaps 0;
 QY 1 RQDLMLFLDSSASVSHYEFSSRVREPVQGLVAPLPGTGLRSLVHVGSRPYTEPPGQH 60
 DB 32 QGDLFLDSSASVSHYEFSSRVREPVQGLVATWSFGGALRSLVHVGSRPYTEPPGQH 91
 QY 61 SSGEAAQDAVRASACMGDTHTGLALVAKGQLFAASGARRPVRYLVWTDGSSDDPV 120
 DB 92 SSGQAIRDAIRVAPQRMGDTHTGLALVAKGQLFAEAGARRPVRYLVWTDGSSDDPV 151
 QY 121 GPPMOELKDLGVTVFVSTGRGNFLFLSAASAPAEKHLHFVVDVLDLHIIVOELRGSILD 180
 DB 152 GPPMOELKDLGVTVFVSTGRGNFLFLSAASAPAEKHLHFVVDVLDLHIIVOELRGSILD 211

RESULT 3
 0923K3 MOUSE PRELIMINARY; PRT; 415 AA.
 ID 0923K3_MOUSE PRELIMINARY; PRT; 415 AA.
 AC 0923K3;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Von Willebrand factor A-related protein.
 GN Name=Vwal; Synonyms=4932416A1Rik, Wafp;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22057805; PubMed=12062410; DOI=10.1016/S0014-5793(02)02579-6;
 RA Fitzgerald J., Ting S.-T., Bateman J.F.;
 RT "WARP a new member of the von Willebrand factor A-domain superfamily
 of extracellular matrix proteins.";
 RL FEBS Lett. 517:61-66(2002).
 DR EMBL; AY030094; AAK8350.1; -; mRNA.
 DR HSSP; P18614; IMHP.
 DR Ensembl; ENSMUSG0000042116; Mus musculus.
 DR MGI; MGI:2179729; 4932416A1Rik.
 DR MGI; MGI:2179729; Vwal.
 DR GO; GO:0005615; Cytoplasmic space; TAS.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR002035; VWF_A.

DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00853; FN3; 2.
 DR PROSITE; PS0234; VWF_A; 1.
 DR PROSITE; PS0234; VWF_A; 1.
 SQ SEQUENCE 415 AA; 44737 MW; C3344AECB3FDE431 CRC64;

Query Match 84.1%; Score 768; DB 2; Length 415;
 Best local Similarity 83.3%; Pred. No. 1.7e-59;
 Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
 QY 1 RQDLMLFLDSSASVSHYEFSSRVREPVQGLVAPLPGTGLRSLVHVGSRPYTEPPGQH 60
 DB 32 QGDLFLDSSASVSHYEFSSRVREPVQGLVATWSFGGALRSLVHVGSRPYTEPPGQH 91
 QY 61 SSGEAAQDAVRASACMGDTHTGLALVAKGQLFAASGARRPVRYLVWTDGSSDDPV 120
 DB 92 SSGQAIRDAIRVAPQRMGDTHTGLALVAKGQLFAEAGARRPVRYLVWTDGSSDDPV 151
 QY 121 GPPMOELKDLGVTVFVSTGRGNFLFLSAASAPAEKHLHFVVDVLDLHIIVOELRGSILD 180
 DB 152 GPPMOELKDLGVTVFVSTGRGNFLFLSAASAPAEKHLHFVVDVLDLHIIVOELRGSILD 211

RESULT 4
 08C007 MOUSE PRELIMINARY; PRT; 415 AA.
 ID 08C007_MOUSE PRELIMINARY; PRT; 415 AA.
 AC 08C007;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
 DE library, clone:4932416A1 product:VON WILLEBRAND FACTOR A-RELATED
 DE PROTEIN homolog.
 GN Name=Vwal; Synonyms=4932416A1Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Komoto H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Asahara M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kociba H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nishida I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guernicich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima Y., Mazzarelli J., Mombarto P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schonbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690 (2001).

OY 121 GPMQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLD 180
 DB 152 GPMQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLD 211

RESULT 6
 Q4SCD1_TETNG PRELIMINARY; PRT; 505 AA.

DT 13-SEP-2005 (TREMBlrel. 31, Created)
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF14659, whole genome shotgun sequence.
 GN ORFNames=GSTENG00020556001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 NC NCB1_TaxID=99883;
 RN [1]

RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Cattolico L., Poulain J., De Bernardinis V.,
 RA Biemont C., Skalli Z., Brottier P., Coutanceau J.P., Guzy J.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Lander V., Schachter V., Queller F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 CC Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: CAE01014659; CAC01701.1; -: Genomic DNA.
 SO SEQUENCE 505 AA; 55316 MW; C724FA399E0751FC CRC64;

Query Match 43.5%; Score 397; DB 2; Length 505;
 Best Local Similarity 44.2%; Pred. No. 1.5e-26;
 Matches 80; Conservative 35; Mismatches 64; Indels 2; Gaps 1;

OY 2 GDMLMLSSASVSHTFRRVFEVGLVAPLPIGTALRASLVHVSRRPYEPFGQHS 61
 DB 37 GDVLLTLSSGSSVSHTFRRVFEVGLVAPLPIGTALRASLVHVSRRPYEPFGQHS 96
 OY 62 SGEAQAQAVRASAQRMQDTHGLALVYAKEOLFSAASAPAEKHLHFVDDDLHIIVQELRGSL 119
 DB 97 TQSSAQGSIRIKTPIRGDNTIVALKVAQERYLRGCVGGAAGAPRYLVWLTLDGSSDP 156
 OY 120 VGPQOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSL 179
 DB 157 ISGPAALREBQVAVLVSTGHSNTLVLRITVTAPEVNTLHFVDDDLHIIVQELRGSL 216
 OY 180 D 180
 DB 217 E 217

RESULT 7
 Q4SD22_TETNG PRELIMINARY; PRT; 1723 AA.
 ID Q4SD22_TETNG PRELIMINARY; PRT; 1723 AA.
 AC Q4SD22_TETNG PRELIMINARY; PRT; 1723 AA.

DT 13-SEP-2005 (TREMBlrel. 31, Created)
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
 DE Chromosome 14 SCAF14645, whole genome shotgun sequence.
 GN ORFNames=GSTENG00020231001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 NC NCB1_TaxID=99883;
 RN [1]

RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Cattolico L., Poulain J., De Bernardinis V.,
 RA Biemont C., Skalli Z., Brottier P., Coutanceau J.P., Guzy J.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Lander V., Schachter V., Queller F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 CC Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: CAE01014645; CAG01460.1; -: Genomic DNA.
 DR InterPro: IPR003961; FN III.
 DR InterPro: IPR003962; FNIII subd.
 DR InterPro: IPR02035; VWF_A.
 DR Pfam: PF00041; fn3; 11.
 DR Pfam: PF00092; VMA; 3.
 DR PRINTS: PR00014; ENTPELII.
 DR PRINTS: PR00453; VWFADOMAIN.
 DR SMART: SM00060; FN3; 11.
 DR SMART: SM00327; VMA; 3.
 DR PROSITE: PS50853; FN3; 12.
 DR PROSITE: PS50234; VMPA; 3.
 DR Collagen; Extracellular matrix; Structural protein.
 FT NON TER 1
 FT NON TER 1723
 SO SEQUENCE 1723 AA; 187496 MW; 7EB78DA80D1E6B40 CRC64;

Query Match 30.0%; Score 273.5; DB 2; Length 1723;
 Best Local Similarity 37.4%; Pred. No. 5.2e-15;
 Matches 67; Conservative 28; Mismatches 83; Indels 1; Gaps 1;

OY 1 RGDMLFLSSASVSHTFRRVFEVGLVAPLPIGTALRASLVHVSRRPYEPFGQHS 60
 DB 436 QADVLLVLDGSSVSHTFRRVFEVGLVAPLPIGTALRASLVHVSRRPYEPFGQHS 495
 OY 61 SGEAQAQAVRASAQRMQDTHGLALVYAKEOLFSAASAPAEKHLHFVDDDLHIIVQELRGSL 120
 DB 496 HNLQAVVTRKTRFPYRGSGSTNGRAMTVRETVFGASGARAHVRVITLIDGSSDP 555
 OY 121 GPMQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSL 178
 DB 556 QEPANILNNSDVEIFAVGVKDAVRSLEAIANAPAEKHLHFVDDDLHIIVQELRGSL 614

RESULT 8
 Q4SX3_TETNG PRELIMINARY; PRT; 2225 AA.
 ID Q4SX3_TETNG PRELIMINARY; PRT; 2225 AA.

AC OASX3;
 DT 13-SEP-2005 (TReMBLrel. 31, Created)
 DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF12445, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG00010895001;
 OS Tetradon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neocleosteii;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 Niclaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 Da Silva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 Anthonard V., Jubin C., Castell V., Katinka M., Vacherie B.,
 Biemont C., Skalli Z., Catcolico L., Poullain J., De Berardinis V.,
 Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 Lander V., Schachter V., Queller F., Saurin W., Searpelli C.,
 Wincker P., Lander E.S., Weissbach J., Roest Croillius H.,
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype."
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope, Whitehead Institute Centre for Genome Research,
 RL Submitted (Feb-2004) to the EMBL/Genbank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; CAEB0102445; CAP94689.1; -; Genomic_DNA.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR003961; FN.II.
 DR InterPro: IPR003129; Laminin G_TSP_N.
 DR InterPro: IPR002020; Lectin_LegB.
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF00041; FN3; 8.
 DR Pfam: PF00092; VWA; 3.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 9.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 3.
 DR PROSITE; PS50853; FN3; 9.
 DR PROSITE; PS50307; LECTIN_LEGUME_BETA; UNKNOWN_1.
 DR PROSITE; PS50234; WMPA; 3.
 KW Collagen; Extracellular matrix; structural protein.
 FT NON_TER
 SQ SEQUENCE 2225 AA; 239674 MW; EC6545B60A778BD9 CRC64;
 Query Match 30.0%; Score 273.5; DB 2; Length 2225;
 Best Local Similarity 36.5%; Pred. No. 6.9e-15;
 Matches 66; Conservative 28; Mismatches 82; Indels 5; Gaps 2;
 QY 3 DLMLFLDSSASVSHYEFSSRVREFFVQLVAPLPGLTGALRASLVHVGSRPYTEFFPGQHS 62
 DB 8 DMVFLVDGSWSIGRTNFRILVRVLESLVKAFFVDVDRIRIGLAQFSGEPRLEWHLNTHTT 67
 QY 63 GEAADAVRASARQRMGDTHITGLALVYAKEQLFAKSGARPGPKLVVWTTGGSSDPVGP 122
 DB 68 KEAVWEAARNLPYKGGNTLTGLALFFILENSFSSPSSGRPGIPKIGVLLTGKSDDDVTP 127
 QY 123 PMOELKDLGVTVFIYSTGRGNFLELSAASAPAEKHLHFVDDDDHI---IYQELRGSI 179
 DB 128 PAQRLRDAGVEVFAIGVKNADGELRALIASVEDTHY--NVADPHLMADIVDVTLRITIC 185

QY 180 D 180
 DB 186 E 186
 RESULT 9
 QARP12_TETNG PRELIMINARY; PRT; 1259 AA.
 ID QARP12_TETNG
 AC QARP12;
 DT 13-SEP-2005 (TReMBLrel. 31, Created)
 DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
 DE Chromosome 10 SCAF15009, whole genome shotgun sequence.
 GN ORFNames=GSTENG00031322001;
 OS Tetradon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neocleosteii;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 Niclaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 Da Silva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 Anthonard V., Jubin C., Castell V., Katinka M., Vacherie B.,
 Biemont C., Skalli Z., Catcolico L., Poullain J., De Berardinis V.,
 Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 Lander V., Schachter V., Queller F., Saurin W., Searpelli C.,
 Wincker P., Lander E.S., Weissbach J., Roest Croillius H.,
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype."
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope, Whitehead Institute Centre for Genome Research,
 RL Submitted (Feb-2004) to the EMBL/Genbank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; CAEB01015009; CAG09870.1; -; Genomic_DNA.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR003961; FN.II.
 DR InterPro: IPR003129; Laminin G_TSP_N.
 DR InterPro: IPR002020; Lectin_LegB.
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF00041; FN3; 8.
 DR Pfam: PF00092; VWA; 3.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 9.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 3.
 DR PROSITE; PS50853; FN3; 9.
 DR PROSITE; PS50307; LECTIN_LEGUME_BETA; UNKNOWN_1.
 DR PROSITE; PS50234; WMPA; 3.
 KW Collagen; Extracellular matrix; structural protein.
 FT NON_TER
 SQ SEQUENCE 1259 AA; 138606 MW; 844C01B6FE2D0E5D CRC64;
 Query Match 29.4%; Score 268.5; DB 2; Length 1259;
 Best Local Similarity 36.3%; Pred. No. 1e-14; Indels 1; Gaps 1;
 Matches 65; Conservative 29; Mismatches 84;
 QY 1 RGDLMLDSSASVSHYEFSSRVREFFVQLVAPLPGLTGALRASLVHVGSRPYTEFFPGQHS 60
 DB 393 QADVLVLVDGSYSIGLQNFARVAFLEVLVNSPFIGSKVQISLVQYSRDPHTPEFALNTH 452
 QY 61 SSGEADAVRASARQRMGDTHITGLALVYAKEQLFAKSGARPGPKLVVWTTGGSSDPV 120
 DB 453 HDINAVVAAYTFPRYGGSTWTGKAMKVKXKIFVASKARQNPVAVVLLITDGSSDSF 512
 QY 121 GPMOELKDLGVTVFIYSTGRGNFLELSAASAPAEKHLHFVDDDDHI---IYQELRGSI 178
 DB 513 KDAATNLNINIDVELFPAVGKDVASSELALANPADDNVFEVDFDAFORLSKELTOSI 571
 RESULT 10
 COEAL_CHICK STANDARD; PRT; 1888 AA.
 ID COEAL_CHICK
 AC P32018; Q6LBU0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Collagen alpha 1(XIV) chain precursor (Undulin).
 GN Name=COL14A1;

OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=93185668; PubMed=8444186;
 RA Maechli C., Trub J., Kessler B., Winterhalter K.H., Trub B.;
 RT "Complete primary structure of chicken collagen XIV.";
 RL Eur. J. Biochem. 212:483-490 (1993).
 RN [2]
 RP NUCLEOTIDE SEQUENCE OF 402-1549.
 RX MEDLINE=92339443; PubMed=1339349;
 RA Trub J., Trub B.;
 RT "Type XIV collagen is a variant of undulin.";
 RL Eur. J. Biochem. 207:549-557 (1992).
 RN [3]
 RP NUCLEOTIDE SEQUENCE OF 1582-1770.
 RX MEDLINE=92037585; PubMed=1935930;
 RA Gordon M., Castagnola P., Dublet B., Linsemayer T.F.;
 RA van der Rest M., Mayne R., Olsen B.R.;
 RT "Cloning of the cDNA for a new member of the class of fibril-associated collagens with interrupted triple helices";
 RL Eur. J. Biochem. 201:333-338 (1991).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 1582-1770.
 RA Apte S.S.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP STRUCTURE BY NMR OF 1853-1885.
 RX MEDLINE=96357967; PubMed=9694594; DOI=10.1016/S0945-053X(98)90027-0;
 RA Giry-Lozignez C., Aubert-Foucher E., Penin F., Deleage G., Dublet B., van der Rest M.;
 RT "Identification and characterization of a heparin binding site within the NC1 domain of chicken collagen XIV.";
 RL Matrix Biol. 17:145-149 (1998).
 RN [6]
 RP STRUCTURE BY NMR OF 1852-1885.
 RX MEDLINE=99280705; PubMed=10350466; DOI=10.1021/bi9900222;
 RA Montserret R., Aubert-Foucher E., McLeish M.J., Hill J.M., Flicheux D., Jaquinod M., van der Rest M., Deleage G., Penin F.;
 RT "Structural analysis of the heparin-binding site of the NC1 domain of collagen XIV by CD and NMR.";
 RL Biochemistry 38:6479-6488 (1999).
 CC -1- FUNCTION: An adhesive role by integrating collagen bundles. It is probably associated with the surface of interstitial collagen fibrils via COL1. The COL2 domain may then serve as a rigid arm which sticks out from the fibril and protrudes the large N-terminal globular domain into the extracellular space, where it might interact with other matrix molecules or cell surface receptors.
 CC -1- SUBUNIT: Homotrimer (Probable).
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Wide tissue distribution; high presence in dense connective tissue in skeletal muscle.
 CC -1- PTM: Lysines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- PTM: May contain numerous cysteine residues involved in inter- and intramolecular disulfide bonding.
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with interrupted helices (FACIT) family.
 CC -1- SIMILARITY: Contains 8 fibronectin type-III domains.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -1- SIMILARITY: Contains 2 VWFA domains.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not

CC removed.
 DR EMBL, X70793; CAA50064.1; -; mRNA.
 DR EMBL, X70792; CAA50063.1; -; mRNA.
 DR EMBL, X66138; CAA46928.2; -; mRNA.
 DR EMBL, X65122; CAA46238.1; -; mRNA.
 DR PIR, A45974; A45974.
 DR PIR, S78476; S78476.
 DR PDB, 1B9P; NMR; A=1853-1885.
 DR PDB, 1B9Q; NMR; A=1853-1885.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR02035; VWFA_A.
 DR Pfam; PF00041; fn3; 8.
 DR Pfam; PF00092; VWFA; 2.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR PROSITE; PS50853; FN3; 8.
 DR PROSITE; PS50234; VWFA; 2.
 DR 3D-structure; Cell adhesion;
 KW Glycoprotein; Hydroxylation;
 FT SIGNAL 1 28
 FT CHAIN 29 1888
 FT DOMAIN 29 117
 FT DOMAIN 158 330
 FT DOMAIN 352 440
 FT DOMAIN 442 532
 FT DOMAIN 534 621
 FT DOMAIN 625 711
 FT DOMAIN 742 830
 FT DOMAIN 832 921
 FT DOMAIN 925 1009
 FT DOMAIN 1042 1215
 FT DOMAIN 1239 1444
 FT REGION 1429 1468
 FT REGION 1469 1620
 FT REGION 1664 1786
 FT MOTIF 1489 1491
 FT MOTIF 1617 1619
 FT COMBINS 721 733
 FT CARBOHYD 138 138
 FT CARBOHYD 1398 1398
 FT CARBOHYD 1855 1856
 FT TURN 1859 1861
 FT TURN 1862 1880
 FT HELIX 1881 1882
 FT TURN 1881 1882
 SQ SEQUENCE 1888 AA; 202668 MW; 39915BB9F46DD873 CRC64;
 Query Match 29.1%; Score 265.5; DB 1; Length 1888;
 Best Local Similarity 34.5%; Pred. No. 2.9e-14;
 Matches 61; Conservative 33; Mismatches 82; Indels 1; Gaps 1;
 QY 3 DLMFLDSSASVSHRESRVREFVQGLVAPLPGLGALRALSLVHVGSRPYTFPPGQHS 62
 DB 158 DIVILVDSWSIGRNFVLVRFLENLVSAFVSGSEKTRVGLAQS GSDPRLEMLNAYGT 217
 QY 63 GEAADAVRASQGRGDTHTGLATVYAKEQLEFAEASGAPGPKVLTWTTDGGSSDPVPG 122
 DB 218 KDVAIDAVRANLPYKKGNTLTGLATYIIENSKPAPGARPGVSKGILITDGSDDVIP 277
 QY 123 PWQELKDLGVTVFVSTGRGFLSLAASAPAEKGLHFV-DVDDLIIIVQELRGS 178
 DB 278 PAKNRDAGIELFALGVKNADINELKEIASBPDSTHYVNVADPFNMSIVEGLTRTV 334
 RESULT 11
 ID COCA1 CHICK STANDARD; PRT; 3124 AA.
 AC P13944; 004509;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor (fibrochimertin).


```
FT COMBIAS 3086 3096 Asp/Glu-rich (acidic).
FT COMBIAS 3111 3123 Arg/Lys-rich (basic).
FT CARBOHYD 32 32 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 797 797 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 890 890 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 981 981 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 981 981 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 1006 1006 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1032 1032 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1044 1044 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1512 1512 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1767 1767 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2210 2210 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2273 2273 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2532 2532 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2683 2683 N-linked (GlcNAc...) (Potential).
FT VARSPIC 25 1188 Missing (in isoform Short).
FT CONFLICT 1258 1258 T -> S (in Ref. 4).
FT CONFLICT 1264 1264 D -> E (in Ref. 4).
FT CONFLICT 2759 2759 P -> A (in Ref. 2).
FT CONFLICT 2803 2803 L -> F (in Ref. 2).
FT CONFLICT 2977 2977 V -> F (in Ref. 2).
FT CONFLICT 3075 3075 OP -> AG (in Ref. 3).
SQ SEQUENCE 3124 AA; 340582 MW; 094285AFB7F346CF CRC64;
```

```
Query Match 29.1%; Score 265.5; DB 1; Length 3124;
Best Local Similarity 34.6%; Pred. No. 5,2e-14;
Matches 62; Conservative 33; Mismatches 83; Indels 1; Gaps 1;
```

```
OY 1 RSDLMFLDSSASVSHYEFVRVEFGOLVAPLPGLTGALRASLVVSGRPYTFEPFGQH 60
DB 437 KADVFLVGVSGSIGIANFVKRAFLEVIVKSEISPRKQVSLVGYSDPHMEFLNXY 496
OY 61 SSGEAAQDAVRASAOQMDTHGLALVAKEQFAPASGARPGVPTLVWTDGSSDPV 120
DB 497 NRVKDIQIAINFEPYSGSTNKGKMTYREKVFVTSKSGSRPVPRVMTLITDGKSDAF 556
OY 121 GPPMDELKGLVTFVTVGRGNFLELSAASAPAKHLHFV-DVDDHLITVQELRGS 178
DB 557 KEPALIKRDADVEIFAVGVKADAVTELEAIASPPATHTVYEDFAPORISFELTQSV 615

RESULT 12
MATN4_MOUSE STANDARD; PRT; 624 AA.
ID MATN4_MOUSE 089029; O9QWS3;
AC 089029; O89030; O9QWS3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Matrilin-4 precursor (Mat-4).
GN Name=Matn4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT), AND VARIANTS.
RC STRAIN=C57BL/6J, and CD-1; TISSUE=Fetal;
RX MEDLINE=98442849; PubMed=9771906; DOI=10.1016/S0014-5793(98)01111-9;
RA Wagner R., Kobbé B., Paulsson M.;
RT "Matrilin-4, a new member of the matrilin family of extracellular
matrix proteins."
RL FEBS Lett. 436:123-127 (1998).
RN [2]
RP NUCLEOTIDE SEQUENCE (LARGE SCALE MRNA) (ISOFORM LONG).
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=23288257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
```

```
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins K.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L.F., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gamarate P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP NUCLEOTIDE SEQUENCE OF 218-257.
RX MEDLINE=99043241; PubMed=9827539; DOI=10.1016/S0014-5793(98)01293-9;
RA Wagner R., Kobbé B., Paulsson M.;
RT "Genomic organization, alternative splicing and primary structure of
human matrilin-4."
RL FEBS Lett. 438:165-170 (1998).
CC -1- FUNCTION: Major component of the extracellular matrix of
cartilage.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
IsoId=O89029-1; Sequence=Displayed;
Name=Short;
IsoId=O89029-2; Sequence=VSP 001401;
CC -1- TISSUE SPECIFICITY: Lung, brain, sternum, kidney and heart.
CC -1- DEVELOPMENTAL STAGE: The short isoform was detected in 7 weeks old
mice but not in developing mice (19.5 dpc embryos or in 2, 8, and
21 days old animals).
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 2 WFPA domains.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC -----
CC EMBL: AJ006140; CA06889.1; -; mRNA.
CC EMBL: AJ006140; CA06890.1; -; mRNA.
CC EMBL: BC036558; AA036558.1; -; mRNA.
CC EMBL: AJ010984; CA09451.1; -; Genomic DNA.
CC HSSP: P00736; IAP0.
CC EMBL: ENSMUSG0000016995; Mus musculus.
CC WGI: WGI:138314; Matn4.
CC GO: GO:0005615; C:extracellular space; TAS.
CC InterPro: IPR000152; Asx hydroxyl_S.
CC InterPro: IPR000742; EGF_2.
CC InterPro: IPR006209; EGF-like.
CC InterPro: IPR011203; Matn_WFA.
CC InterPro: IPR02035; WFA_A.
CC Pfam: PF00008; EGF_4.
CC Pfam: PF00092; WFA; 2.
CC PIRSF: PIRSF015217; Matn_WFA; 1.
CC PRINTS: PR00453; WFAFADOMAIN.
CC PROSITE: PS00010; ASX HYDROXYL; 2.
CC PROSITE: PS01186; EGF_2; 2.
CC PROSITE: PS50026; EGF_3; FALSE_NEG.
CC PROSITE: PS50234; WFA; 2.
CC KX Polymorphism; Repeat; Signal.
CC SIGNAL 1 21 Potential.
CC CHAIN 22 624 Matrilin-4.
CC DOMAIN 36 215 WFA 1.
```


FT DOMAIN 217 257 BGF-like 1.
 FT DOMAIN 258 298 EGF-like 2.
 FT DOMAIN 299 339 EGF-like 3.
 FT DOMAIN 340 380 EGF-like 4.
 FT DOMAIN 388 563 WMFA 2.
 FT COILED 590 623 Potential.
 FT CARBOXD 71 71 N-linked (GlcNAc. . .) (Potential).
 FT CARBOXD 307 307 N-linked (GlcNAc. . .) (Potential).
 FT DISULFD 221 232 By similarity.
 FT DISULFD 228 241 By similarity.
 FT DISULFD 243 256 By similarity.
 FT DISULFD 262 273 By similarity.
 FT DISULFD 269 282 By similarity.
 FT DISULFD 284 297 By similarity.
 FT DISULFD 303 314 By similarity.
 FT DISULFD 310 323 By similarity.
 FT DISULFD 325 338 By similarity.
 FT DISULFD 344 355 By similarity.
 FT DISULFD 351 364 By similarity.
 FT DISULFD 366 379 By similarity.
 FT VARSPLC 28 217 Missing (in isoform short).
 FT VARIANT 319 319 /FTid=VSP_001401.
 FT VARIANT 346 346 A -> E (in strain C57BL/6J).
 FT SEQUENCE 624 AA; 68918 MW; DPA28D2C94B1A14F CRC64;

Query Match 29.0%; Score 265; DB 1; Length 624;
 Best Local Similarity 36.8%; Pred. No. 9.2e-15;
 Matches 64; Conservative 29; Mismatches 75; Indels 6; Gaps 2;

Qy 3 DLMLFLSSASVSHYFSRVRFPVGVLPPLGTGALRASLVHVSRRPYTEFPFGQSS 62
 Db 36 DLVIMIDSRSRYPRPEFTMRQFLVGLRLSDVGINATRVGVIOYQSQVSFPLGAFSR 95
 Qy 63 GEAQADAVRASQNRGDTHTGLALVYAKQQLFAEASGARPG--VPKYLWVTTDGSDDP 119
 Db 96 REDMERAIRAVVPLAQTMTGLAIQYAMNVAFSEEGARPSERPRVIVITDGRPDQR 155
 Qy 120 VGPPMQLKDLGVTVFIVSTGRGNFLSLAASAPAEKHLHFVDVDDHLIIYQEE 173
 Db 156 VAEVAQAQRARGIEIYAVGVQADVGSRLRTMASPLDQHVLFVESFDL---IQE 206

RESULT 13
 QASH63.TETNG PRELIMINARY; PRT; 1557 AA.
 ID QASH63.TETNG
 AC QASH63
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome 8 SCAR14587, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG00018311001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OC NCB1_TaxID=99983;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Maucell E., Bounaceau L., Fischer C., Ozouf-Costat C., Bernot A.,
 RA Nicaut S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dastay C., Salanoubat M., Levy M., Boudet N., Cascellano S.,
 RA Anhouard V., Jubin C., Gaetelli V., Katinka M., Vacherie B.,
 RA Bismont C., Skalli Z., Cateolico L., Poulain J., De Berardinis V.,
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Lauder V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;

RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RG Genoscope, Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.

DR EMBL: CAAB0104587; CAG00019.1; -; Genomic_DNA.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR003961; FN III.
 DR InterPro: IPR003129; Laminin_G_TSP_N.
 DR InterPro: IPR02035; VWF_A.
 DR Pfam: PF01391; Collagen_2.
 DR Pfam: PF00041; FN3; 6.
 DR Pfam: PF00092; VWA; 2.
 DR PRINTS: PR00453; VWFADOMAIN.
 DR SMART: SM00060; FN3; 6.
 DR SMART: SM00210; TSPN; 1.
 DR SMART: SM00327; VWA; 2.
 DR PROSITE: PS50853; FN3; 7.
 DR PROSITE: PS50234; VWF; 2.
 DR Collagen; Extracellular matrix; Structural protein.
 FT NON_TER 1
 FT SEQUENCE 1557 AA; 168266 MW; 6203E9A0900C6D9D CRC64;

Query Match 29.0%; Score 265; DB 2; Length 1557;
 Best Local Similarity 34.7%; Pred. No. 2.6e-14;
 Matches 61; Conservative 38; Mismatches 75; Indels 2; Gaps 2;

Qy 1 RGDLMLFLSSASVSHYFSRVRFPVGVLPPLGTGALRASLVHVSRRPYTEFPFGQ 59
 Db 889 KADLVFLVDGWSIGDENFMKTRFLHSTVGSLLDIGNQVAILAQSDDAETEFQLSS 948
 Qy 60 HSSEGAADAVRASQNRGDTHTGLALVYAKQQLFAEASGARPGVPKYLWVTTDGSDDP 119
 Db 949 HSNKEALLAEAIQKISYKGNKTKRAIKGVESIFSLFAGARRGVPKYLVLTDGRSDD 1008
 Qy 120 VGPPMQLKDLGVTVFIVSTGRGNFLSLAASAPAEKHLHFV-DVDDHLIIYQEE 174
 Db 1009 VNKYSKEMQMDGYIIFAIIGFADADYGLVNIASKPSDHHVFFVDDLDVAKLIEQL 1064

RESULT 14
 OSVYK2.HUMAN PRELIMINARY; PRT; 2884 AA.
 ID OSVYK2.HUMAN
 AC OSVYK2
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Collagen, type XII, alpha 1.
 GN Name=COL12A1; ORFNames=RP1-238D15.1-003;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OC NCB1_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Tirmann A.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Smith M.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Corby N.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL354664; CAH171309.1; -; Genomic_DNA.
 DR EMBL: AL080250; CAL19897.1; -; Genomic_DNA.


```
FT DOMAIN 1655 1652 Fibronectin type-III 11.
FT DOMAIN 1654 1743 Fibronectin type-III 12.
FT DOMAIN 1752 1841 Fibronectin type-III 13.
FT DOMAIN 1843 1931 Fibronectin type-III 14.
FT DOMAIN 1933 2022 Fibronectin type-III 15.
FT DOMAIN 2024 2113 Fibronectin type-III 16.
FT DOMAIN 2115 2202 Fibronectin type-III 17.
FT DOMAIN 2206 2290 Fibronectin type-III 18.
FT DOMAIN 2323 2496 TSP N-terminal.
FT DOMAIN 2520 2712 Nonhelical region (NC3).
FT REGION 2451 2746 Triple-helical region (COL2) with 1
FT REGION 2747 2898 imperfection.
FT REGION 2899 2941 Nonhelical region (NC2).
FT REGION 2942 3044 Triple-helical region (COL1) with 2
FT REGION 3045 3063 imperfections.
FT MOTIF 862 864 Cell attachment site (Potential).
FT MOTIF 2779 2781 Cell attachment site (Potential).
FT MOTIF 2895 2897 Cell attachment site (Potential).
FT MOD_RES 2944 2944 Hydroxyproline (By similarity).
FT MOD_RES 2947 2947 Hydroxyproline (By similarity).
FT MOD_RES 2950 2950 Hydroxyproline (By similarity).
FT MOD_RES 2959 2959 Hydroxyproline (By similarity).
FT MOD_RES 2965 2965 Hydroxyproline (By similarity).
FT MOD_RES 2968 2968 Hydroxyproline (By similarity).
FT MOD_RES 2971 2971 Hydroxyproline (By similarity).
FT MOD_RES 2983 2983 Hydroxyproline (By similarity).
FT MOD_RES 3000 3000 Hydroxyproline (By similarity).
FT MOD_RES 3003 3003 Hydroxyproline (By similarity).
FT MOD_RES 3014 3014 Hydroxyproline (By similarity).
FT MOD_RES 3023 3023 Hydroxyproline (By similarity).
FT MOD_RES 3026 3026 Hydroxyproline (By similarity).
FT MOD_RES 3029 3029 Hydroxyproline (By similarity).
FT CARBOHYD 700 700 O-linked (Xyl...) (Potential).
FT CARBOHYD 798 798 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 889 889 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 981 981 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 1763 1763 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2206 2206 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2528 2528 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2679 2679 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 25 1188 Missing (in isoform Short).
SQ SEQUENCE 3063 AA; 333194 MW; 75FBA78FA8E48293 CRC64;

Query Match 28.9%; Score 263.5; DB 1; Length 3063;
Best Local Similarity 35.8%; Pred. No. 7.7e-14;
Matches 64; Conservative 30; Mismatches 84; Indels 1; Gaps 1;
```

Search completed: February 13, 2006, 07:47:45
Job time : 144.074 secs

PN WO200288184-A1.
 XX 07-NOV-2002.
 XX
 XX
 PE 02-MAY-2002; 2002MO-AU000542.
 XX
 XX
 PR 02-MAY-2001; 2001AU-00004701.
 XX
 XX
 PA (MURD-) MURDOCH CHILDRENS RES INST.
 XX
 XX Bateman JF, Fitzgerald DJ;
 DR WPI; 2003-111873/10.
 XX N-PSDB; AAD50399.
 XX
 PT New isolated Willebrand Factor A-related Protein polypeptide useful for
 PT the manufacture of a medicament in the treatment of a disease condition
 PT of the extracellular matrix, in particular arthritis.
 XX
 PS Claim 11; Page 76-78; 103pp; English.

The invention relates to Willebrand Factor A domain related-protein
 (WARP) which is a member of von Willebrand Factor A (vW)-domain protein
 superfamily of extracellular matrix (ECM) proteins. WARP is used as a
 molecular marker, used for detecting a loss of ECM integrity in an animal
 subject, monitoring repair, regeneration or other disease processes in an
 animal subject and detecting a disease condition or a propensity for the
 development of a disease condition in an animal subject. The invention is
 useful for the manufacture of a medicament in the treatment of a disease
 condition of the ECM. The disease condition involves the cartilage, and
 is preferably arthritis. The invention is also used in gene therapy. The
 present sequence is human WARP protein

Sequence 418 AA;

Alignment Scores:

Pred. No.: 1,88e-138 Length: 418
 Score: 2137.50 Matches: 417
 Percent Similarity: 99.8% Conservative: 0
 Best Local Similarity: 99.8% Mismatches: 0
 Query Match: 89.8% Indels: 1
 DB: 6 Gaps: 1

US-10-699-035A-5 (1-1254) x AAE32502 (1-418)

QY 1 ATGCTCCCTGGAGCGGCTGGGCTGAGCTTGCGGCTGGCGCGAGC 60
 DB 1 MetLeuProTrpThrAlaLeuGlyLeuAlaLeuSerLeuGlyLeuAlaArgSer 20
 QY 61 GCGCGGAGCGGCTGCCAGCATCAGCCCGCGAGGAGCTGATGTTCTGCTGAC 120
 DB 21 GlyAlaGlyArgGlyProProAlaSerAlaProArgGlyAspLeuSerPheLeuLeuAsp 40
 QY 121 AGCTCAGCCAGGCTCTCTCACTACAGATTCTCCGGGCTTGGGAGTTTGTGGGACGCTG 180
 DB 41 SerSerAlaSerValSerHisTrpGluPheSerArgValArgGluPheValGlyGluLeu 60
 QY 181 GTGGCTCCAGTCCCTGGGAGCGGCGGCTGGCGGCTGAGTGTGACGCGGAGCT 240
 DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
 QY 241 CGGCGCATACCGAGTTCCTTCGCGCAGCAGAGCTCGGTGAGGCTCCGAGATGCG 300
 DB 81 ArgProTrpThrGluPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla 100
 QY 301 GTGCGTGTCTTGGCCAGCGCATGGGTGACACCCACACATGCGCTGGCGCTGTATGCC 360
 DB 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisTrpGlyLeuAlaLeuValTyrAla 120
 QY 361 AAGGAACAGCTGTTTGTGAGCATAGGTCGCGGCGCGAGGAGTGGCCAAAGTGGTGGT 420
 DB 121 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 140

QY 421 TGGGTACAGATGGCGGCTCCAGCGAGCCCTGTGGGCCCCCCCATACGAGAGCTCAAGGAC 480
 DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGluLeuLysSer 160
 QY 481 CTGGGCGCTCAGCGTTCATTGTGACAGCGCGGAGGCACTTCCTGGAGCTGTCAAGC 540
 DB 161 LeuGlyValThrValPheLeuAlaSerThrGlyArgGlyAsnPheLeuGlyLeuSerAla 180
 QY 541 GCTGCTCAGCCCTGCGGAGAGACACTTGACCTTTGACGTGTGACCTGCACATC 600
 DB 181 AlaAlaSerAlaProAlaGlnGlyLeuHisPheValAspValAspAspLeuHisLe 200
 QY 601 ATTGTCCAAGCTGAGGGGCTCCATTTC--GGATGGCGGCGGACAGCTCATGCC 657
 DB 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
 QY 658 ACGGAGTACAGTCCAGCGGCTTCGCGCTGGCGCTGGGACCCCTGTCAGCGGAGACTCG 717
 DB 221 ThrGluIleTrpSerSerGlyPheArgLeuAlaTrpProProLeuLeuThrAlaAspSer 240
 QY 718 GGTACTATGTGCTGAGAGCTGTGCTCCAGCGCCGAGCGGAGGCTGCAAGACGCGCAG 777
 DB 241 GlyTyrTrpValIleGluLeuValProSerAlaGlnProGlyAlaAlaArgGlnGln 260
 QY 778 CTGCCAGGAAACGCCAGCACTGATCTGGGCGGCGCTGACCCGAGACAGACTACGAC 837
 DB 261 LeuProGlyAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrAsp 280
 QY 838 GTGGGCTCAGTCCGCTGAGTCCAGAGTCCCTCTGAGAGCCCGAGTCTGGGTGGC 897
 DB 281 ValAlaLeuValProGluSerAsnAlaArgLeuLeuAspProGlnIleLeuArgValArg 300
 QY 898 ACGCGCCAGAGAGGCGGCGGCGAGCGCATGCTCATCTCCAGCGCGCGCGCGCAGC 957
 DB 301 ThrArgProGluGlnGluAlaGlyProGluArgIleValIleSerHisAlaArgProArgSer 320
 QY 958 CTCCGCTGAGTGGGCGCCAGCGCTGGGCTGACGCGCGGCGCTGCTACGAGTGCAG 1017
 DB 321 LeuArgValSerTrpAlaProAlaLeuGlySerAlaAlaAlaLeuGlyTyrHisValGln 340
 QY 1018 TTGGGCGGCTGGGCGGCGGAGGCGGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGG 1077
 DB 341 PheGlyProLeuValArgGlyGlyGlnAlaGlnArgValGluValProAlaGlyArgAsnLys 360
 QY 1078 ACCAGCTGACAGGCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 1137
 DB 361 ThrThrLeuGlnGlnGlyLeuAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPhe 380
 QY 1138 CGCTCGGCGCGGAGAGCGCGCTGTCCGCGCAAGGCTGCAAGCGCGCGCGCGCGCG 1197
 DB 381 ArgSerGlyArgGluSerAlaLeuSerAlaLysAlaCysThrProAspGlyProArgPro 400
 QY 1198 CGGCGACCGCGCGTGGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1251
 DB 401 ArgProArgProValProArgAlaProThrProGlyThrAlaSerArgGluPro 418
 RESULT 2
 ID ABP69674 standard; protein; 445 AA.
 AC ABP69674;
 XX
 DT 20-JAN-2003 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 1721.
 XX
 KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
 KW arthritis; cytostatic; immunomodulator; nocotropic; neuroprotective;
 KW antiparkinsonian; antidabetic; immunosuppressive; dermatological;
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

antiarthritic.
Homo sapiens.
MO200270539-A2.
12-SEP-2002.
05-MAR-2002; 2002WO-US005095.
05-MAR-2001; 2001US-00799451.
(HYSE-) HYSEQ INC.
Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F, Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M, Wehman T, Wang J, Wang D, Drmanac RT;
WPI; 2002-759812/82.
N-PSDB; AB211891.
New polynucleotides comprising sequences assembled from expressed sequence tags (ESTs), useful for treating cell-proliferative, neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet or coagulation disorders.
Claim 9; SEQ ID NO 1721; 1012bp + Sequence Listing; English.
The invention relates to an isolated polynucleotide (I) comprising a nucleotide sequence selected from any of 948 sequences (AB211119-AB212066) or their mature protein coding portion, active domain coding protein or complementary sequences. The polynucleotides are useful for identifying expressed genes or for physical mapping of human genome. The encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight markers, as a food supplement, for generating antibodies, in medical imaging, screening and diagnostic assays and for treating cell-proliferative disorders (cancer), neurodegenerative diseases (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid disorders, CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal, parasitic), arthritis, etc. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pcl_sequences
SQ Sequence 445 AA;
Alignment Scores:
Pred. No.: 7,85e-137 Length: 445
Score: 2114.00 Matches: 417
Percent Similarity: 93.7% Conservative: 0
Best Local Similarity: 93.7% Mismatches: 0
Query Match: 88.8% Indels: 28
DB: 5 Gaps: 2
US-10-699-035A-5 (1-1254) x ABP69674 (1-445)
QY 1 AAGCTCCCTGACGGCGCTCGGCGCTGAGCTTGGCGCGCTGGCGCGGAGC 60
DB 1 MetLeuProTirPThrLLeuGlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
QY 61 GCGCGGAGCGCGGTCACAGCATCAGCCCGCGAGGAGCACTGATGTTCTGCTGAGC 120
DB 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetHeuLeuAsp 40
QY 121 AGCTGACGAGCGCTCTCTCACTAGAGTTCTCCCGGCTTGGAGTTTGTGGGAGCTG 180
DB 41 SerSerAlaSerAlaSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACTGCTGGGCAACCGGGCGCTGGCGGCACTGCTGCACTGGGCACT 240
DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80

QY 241 CGGCATACACCGAGATTCCCTTCCGCGACGACAGCTGGGCTGAGGCTGCCAGATGCG 300
DB 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 100
QY 301 GTGGCGCTTCGCGCGGCGCATGGGTGACACCCACATCGGCGCTGGCGCTGCTGCGC 360
DB 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisSerHisGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGAACAGCTGTTTGTCTGAACATCAGGTGCGCGCGCGAGGAGTGCCTGCTGCTG 420
DB 121 TysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyValProTyrValLeuVal 140
QY 421 TGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCTCCCGCATGACGAGATCAAGAC 480
DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGlnLeuLysAsp 160
QY 481 GTGGGCTTCACCGTGTCTTGTGACGACCGCGCGGCACTTCTTGTGAGCTGTGAGCC 540
DB 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGlnLeuSerAla 180
QY 541 GCTGCTCAGCCCTGCGCGAGAGACACTGACCTTGTGAGCTGATGATGACCTGACATC 600
DB 181 AlaAlaSerAlaProAlaGluLysHisSerHisPheValAspValAspLeuHisIle 200
QY 601 ATTGTCAGAGCTGAGGGGCTTCATTCTC---GCGATGCGCGCGAGAGCTCCATGCC 657
DB 201 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
QY 658 ACGGAGATACGTCACGCGGCTTCCGCGCTGGCTGGCCACCTCTGTACCGGACCTCG 717
DB 221 ThrGluLeuThrSerSerIlePheArgLeuAlaTrpProLeuLeuThrAlaSer 240
QY 718 GCTACTATGCTGTGAGCTGTGTCGCGCGCGCGCGCGGAGGAGCTGCAAGACCGCAGCAG 777
DB 241 GlyTyrTyrValLeuGlnLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
QY 778 GTGCCAGGAGACCCACGAGCTGATCTGGCGCGGCTGACCCGCGACACGACTACGAC 837
DB 261 LeuProGlyAlaAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrAsp 280
QY 838 GTGGCGCTATGCTGTGAGCTGACGTCACGCGCGCTCTGAGGCGCGCGGCTGCGGCGCGC 897
DB 281 ValAlaLeuValProGluSerAsnValArgLeuLeuArgProGlnIleLeuArgValArg 300
QY 898 ACGCGG----- 903
DB 301 ThrArgProGlyGluAlaGlyProGlyAlaSerGlyProGluSerGlyAlaGlyProAla 320
QY 904 -----CCAGAGAGGCGCGGCGCAGAGCGCATCGTCATC 936
DB 321 ProThrGlnLeuAlaAlaLeuProAlaProAlaProGluGlnAlaGlyProGluArgIleValIle 340
QY 937 TCCACGCGCGCGCGCGAGCTCCGCGTGAAGTTGGGCGCGCGCTGGGCTGACGCGCG 996
DB 341 SerHisAlaArgProArgSerLeuArgValSerTrpAlaProAlaLeuAlaSerAlaAla 360
QY 997 GCGCTCGGCTACAGGTGACGTTCCGCGCGCGCTGCGCGGCGGAGGCGCGAGCGGAG 1056
DB 361 AlaLeuGlyTyrHisValGlnPheGlyProLeuAspGlyGlyGlnAlaGlnArgValGln 380
QY 1057 GTGCCCGCGCGGCGCACTGACACGCTGACGAGGCTTGGCGCGCGGCGCGCTACCTG 1116
DB 381 ValProAlaGlyArgAsnGlySerThrThrLeuGlnGlyLeuAlaProGlyThrAlaTyrLeu 400
QY 1117 GTGACCGTGAACCGCGCTTCCGCTCGGCGCGCGAGAGCGCGCTGTCCGCGCAAGGCTGC 1176
DB 401 ValThrValThrAlaAlaPheArgSerGlyArgGluSerAlaLeuSerAlaLysAlaCys 420
QY 1177 ACGCCGAGAGCGGAGC 1236
DB 421 ThrProAspGlyProArgProArgProArgProValProArgAlaProThrProGlyThr 440
QY 1237 GCCAGCGCTGAGCGC 1251

```

Db          441 AlaserArgClnPro 445
RESULT 3
ADH71106
XX          ADH71106 standard; protein; 445 AA.
XX          ADH71106;
XX          25-MAR-2004 (first entry)
XX          Human protein of the invention NOVA SEQ ID NO.2.
XX          human; cytostatic; immunomodulator; neuroprotective; nootropic;
XX          anorectic; antidiabetic; antimicrobial; antidiabetic; gene therapy;
XX          vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
XX          obesity; diabetes; infectious disease; metabolic syndrome X;
XX          dyslipidaemia.
XX          Homo sapiens.
XX          OS
XX          WO2003102155-A2.
XX          11-DEC-2003.
XX          03-JUN-2003; 2003WO-US017430.
XX          03-JUN-2002; 2002US-0385120P.
XX          04-JUN-2002; 2002US-0385784P.
XX          05-JUN-2002; 2002US-0386041P.
XX          06-JUN-2002; 2002US-0386047P.
XX          06-JUN-2002; 2002US-0386376P.
XX          06-JUN-2002; 2002US-0386453P.
XX          06-JUN-2002; 2002US-0386643P.
XX          06-JUN-2002; 2002US-0387016P.
XX          07-JUN-2002; 2002US-0386796P.
XX          07-JUN-2002; 2002US-0386816P.
XX          07-JUN-2002; 2002US-0386931P.
XX          07-JUN-2002; 2002US-0386942P.
XX          07-JUN-2002; 2002US-0386971P.
XX          07-JUN-2002; 2002US-0387262P.
XX          08-JUN-2002; 2002US-0296960P.
XX          10-JUN-2002; 2002US-0387400P.
XX          10-JUN-2002; 2002US-0387535P.
XX          11-JUN-2002; 2002US-0387610P.
XX          11-JUN-2002; 2002US-0387625P.
XX          11-JUN-2002; 2002US-0387634P.
XX          11-JUN-2002; 2002US-0387668P.
XX          11-JUN-2002; 2002US-0387696P.
XX          11-JUN-2002; 2002US-0387702P.
XX          11-JUN-2002; 2002US-0387836P.
XX          11-JUN-2002; 2002US-0387859P.
XX          12-JUN-2002; 2002US-0387933P.
XX          12-JUN-2002; 2002US-0387934P.
XX          12-JUN-2002; 2002US-0387960P.
XX          12-JUN-2002; 2002US-0388022P.
XX          12-JUN-2002; 2002US-0388096P.
XX          13-JUN-2002; 2002US-0389123P.
XX          14-JUN-2002; 2002US-0389118P.
XX          14-JUN-2002; 2002US-0389120P.
XX          14-JUN-2002; 2002US-0389144P.
XX          14-JUN-2002; 2002US-0389146P.
XX          17-JUN-2002; 2002US-0389129P.
XX          17-JUN-2002; 2002US-0389142P.
XX          18-JUN-2002; 2002US-0389884P.
XX          19-JUN-2002; 2002US-0390066P.
XX          19-JUN-2002; 2002US-0390209P.
XX          21-JUN-2002; 2002US-0390763P.
XX          17-JUL-2002; 2002US-0396706P.
XX          06-AUG-2002; 2002US-0401628P.
XX          09-AUG-2002; 2002US-0402156P.
XX          09-AUG-2002; 2002US-0402256P.
XX          09-AUG-2002; 2002US-0402389P.

```

```

PR 12-AUG-2002; 2002US-0402786P.
PR 12-AUG-2002; 2002US-0402816P.
PR 12-AUG-2002; 2002US-0402821P.
PR 12-AUG-2002; 2002US-0402832P.
PR 13-AUG-2002; 2002US-0403448P.
PR 13-AUG-2002; 2002US-0403459P.
PR 13-AUG-2002; 2002US-0403531P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0403563P.
PR 13-AUG-2002; 2002US-0406317P.
PR 15-AUG-2002; 2002US-0406317P.
PR 26-AUG-2002; 2002US-0406182P.
PR 26-AUG-2002; 2002US-0406355P.
PR 27-AUG-2002; 2002US-0406240P.
PR 12-SEP-2002; 2002US-0410084P.
PR 20-SEP-2002; 2002US-0412528P.
PR 23-SEP-2002; 2002US-0412731P.
PR 30-SEP-2002; 2002US-0414801P.
PR 30-SEP-2002; 2002US-0414839P.
PR 30-SEP-2002; 2002US-0414840P.
PR 30-SEP-2002; 2002US-0414954P.
PR 09-OCT-2002; 2002US-0417166P.
PR 09-OCT-2002; 2002US-0417406P.
PR 23-OCT-2002; 2002US-0420639P.
PR 28-OCT-2002; 2002US-0421156P.
PR 31-OCT-2002; 2002US-0422690P.
PR 01-NOV-2002; 2002US-0423130P.
PR 05-NOV-2002; 2002US-0423798P.
PR 05-NOV-2002; 2002US-0423798P.
PR 12-NOV-2002; 2002US-0425453P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Alsebrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;
XX Catereron E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
XX Eutenberg S, Gangolli BA, Gerlach VL, Gorman L, Gunther E, Guo X;
XX Gusev VY, Hermann JL, Ji W, Kekuda R, Li L, Liu X, Macdougall JR;
XX Maciachian T, Malyanfarj UM, Mezick AJ, Millet I, Mishra VS;
XX Padigar M, Patnayan M, Pena CE, Peyman JA, Raha D, Raetelli L;
XX Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
XX Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;
XX Zhong H;
XX
XX WPI; 2004-081935/08.
XX
XX N-PesDB; ADH71105.
XX
XX
XX New NOVA polypeptides and nucleic acid molecules useful for preventing or
XX treating NOVA-associated disorders, e.g. cancer, diabetes, infection or
XX obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
XX
XX Example 1; SEQ ID NO 2; 1880bp; English.
XX
XX The invention relates to a novel isolated polypeptide (NOVA) . A
XX polypeptide of the invention has cytostatic, immunomodulator,
XX neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
XX antidiabetic activity, and may have a use in gene therapy, and as a
XX vaccine. The polypeptides are encoded by NOVA polynucleotides comprising
XX any of the 303 fully defined nucleotide sequences given in the
XX specification. The polypeptide is useful in the manufacture of a
XX medicament for treating a syndrome associated with a human disease. The
XX polypeptide, polynucleotide and antibody are useful in diagnosing,
XX treating or preventing NOVA-associated disorders, e.g. cancer, cachexia,
XX Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
XX diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
XX further used as hybridisation probes, in chromosome mapping, tissue
XX typing, preventive medicine, and pharmacogenomics. The present sequence
XX represents a NOVA polypeptide of the invention.
XX
XX
XX Sequence 445 AA;
XX
XX
XX Alignment Scores: 7,85e-137 Length: 445
XX Pred. No.: 2114,00 Matches: 417
XX Score:

```


Kidney injury-associated molecule, KIM, polypeptides - upregulated in injured or regenerating tissues, useful to promote tissue growth and regeneration, especially to treat renal conditions.

The present sequence represents a kidney injury associated molecule (KIM) protein. KIM proteins can be administered therapeutically by expressing KIM encoding polynucleotides, to promote growth and/or survival of damaged tissue (e.g. renal tissue), since the KIM proteins are upregulated in injured or regenerating (especially renal) tissues. KIM fusion proteins, conjugates, antibodies and vectors can also be used therapeutically, e.g. these or the KIM proteins may be included with an acceptable carrier in pharmaceutical compositions, useful for therapy/prophylaxis of conditions associated with dysfunction/dysregulation of KIM genes or proteins, especially renal diseases or impairments of renal function in humans (e.g. acute renal failure, acute nephritis). The polynucleotides can be used to produce antisense sequences which, when internalised into cells, can disrupt expression of a cellular KIM gene, also useful in therapy (e.g. to block the growth of tumours dependent on KIM for growth) or compositions. The proteins and polynucleotides are useful diagnostically e.g. to detect and quantify renal injury/disease (indicative of increased risk, or presence of, renal injury or impaired function), or abnormal responses to tissue injury (indicative of increased risk, or presence of, an autoimmune response or abnormal tissue growth arising from/affecting renal tissue). The proteins can also be used to locate KIM-producing cells (especially specific loci, e.g. tissue masses abnormally producing/expressing KIM such as tumours arising from/affecting renal tissue), by contacting cells with an imageable KIM-binding reagent and imaging reagent accumulation.

SQ Sequence 421 AA;

Alignment Scores:

Pred. No.:	2.34e-103	Length:	421
Score:	1.627.00	Matches:	324
Percent Similarity:	84.9%	Conservative:	31
Best Local Similarity:	77.5%	Mismatches:	59
Query Match:	68.4%	Indels:	
DB:	2	Gaps:	2

US-10-699-035A-5 (1-1254) x AAW86326 (1-421)

OY		1	ATGTCCTCCCTGGACCGCGCCTGGCGGCCGTAGCTTGCAGCTGGCGCTGGGCGCGAGC	60
Db		7	MettLeuPheTrpThValIleuSerMetAlaIeuSerIleuArgIleuAlaIeuAlaIeuSer	26
OY		61	GCGCGGAGCGCGGTRCCAACACATAAGCCCCCAAGGGGAACCTGATGTTCCGTGGAC	120
Db		27	GlyIleGluArgGlyProThrAlaSerAlaProGlnGlyAspIleuPheIleuAsp	46
OY		121	AGCTCAGCGACGGTGCTCTCACTACAGATTCTCCGGGTTCCGGAGATTGTGGGGCAGCTG	180
Db		47	SerSerAlaSerValSerHisTryGluPheSerArgValArgGluPheValGlyGlnIleu	66
OY		181	GTGCTCCACTGCCCCCTGGGACAACCGGGGCGCTGCTGCCAGTTGGTGTCACGTGGCACAGT	240
Db		67	ValAlaThrMetProPheGlyProGlyAlaIeuArgAlaSerIleuValHisValGlySer	86
OY		241	CGGCCATACACCAGAGTTCCTCCCTCGGCGACAGCATCGGTGAGGCTGGCCAGATGCG	300
Db		87	ArgProHisIleThrGluPheThrPheArgGlnTrySerSerGlyAlaIleAlaIleGlnAraIa	106
OY		301	GTGCGTGTCTCTGCCACGCGCATGGGTGAACCCACACATGGCCCTGGCGCTGTCTATGCC	360
Db		107	ValArgValAlaProGlnArgMetGlyAraThrAsnThrGlyIleuAlaIeuAlaTyzAla	126
OY		361	AAGAACAAGCTGTGTTGCTAAAGCATAGGTGGCCGGCGCAAGGGTGCCCAAAGTCTGGTG	420
Db		127	LysGlnGlnIleuPheAlaGlnGlnAlaArgIleuGlyAlaProLysValIleuVal	146
OY		421	TGGGTGACAGATGGCGGCTTCACAGCAACCTGTGGGCCCCCCCATTGCAGAGACTCAAGAC	480

Db	147	TriValThrArgGlyAlaSerSerAspSerValGlyProPrometGlnGluLeuLysAsp	166
OY	481	CTGGGCGCTCACCGTGTTTCATTGTTCAGACACCGCGCCAGAGCACTTCTGGAGCTGTCAACC	540
Db	167	LeuGlyValThrIlePheIleValSerThrGlyArgGlyAsnLeuLeuGluLeuVal	186
OY	541	GGTGGCCACGCCCCCGCGAGAGACCTTCGACTTTGTGGACGGGATGACCTTGACATC	600
Db	187	AlaIaSerAlaProAlaGluLysHISLeuHisPheValAspValAspAspLeuProIle	206
OY	601	ATTGTCCAAAGAGCTGAGGGGCTCCATTCTC---GCGATGGCGCGCAGAGCTCCATGCC	657
Db	207	IleValArgGluLeuLeuArgGlyAlaIleIleAspAlaMetClnProHisGlnLeuHisAla	226
OY	658	ACGGAGATCAAGTCCAGCGGCTTCCGCTGAGCTGAGCCACCCCTGTGACCGCAGACTCG	717
Db	227	SerGluLeuSerSerAsnGlyPheArgLeuSerThrProProLeuLeuThrAlaAspSer	246
OY	718	GGCTACATATGTCTGGAGCTGTGTGCCACAGCGCCACGCGGGGGCTGCAAAACCGCAGAG	777
Db	247	GlyTyrTyrValLeuGluLeuValProSerGlyLysLeuValAlaIleThrArgArgGlnGln	266
OY	778	CTGCGCAGGAAACGCCACGCACTGGATGTGGCGCGGCTCCAGCCCGGACACGCACTACAC	837
Db	267	LeuProGlyAsnAlaThrSerThrThrThrThrAspLeuAsnProAspThrAspTyrGlu	286
OY	838	GTTGGCGCTAGTGCCTGAGTCCAAAGTGCAGCTTCCTGAGGCGCCAGATCTTGGCGGTGCG	897
Db	287	ValSerLeuLeuProGlnSerAsnValArgLeuLeuArgProGlnHisLeuArgValArg	306
OY	898	ACCGGGCCAGAGAGGCGGGGCGCCAGAGCGGATGTACTTCCACGCCCGCGCGGAGC	957
Db	307	ThrLeuGlnGlnGluAlaGlyProGlnArgIleValIleSerHisThrArgProArgSer	326
OY	958	CTCCGCGGTGATGGGCGCCAGCGCTGGGCTCAGCGCGCGCTCGGCTACCAAGTGCAG	1017
Db	327	LeuArgValSerThrAlaProAlaLeuGlyProAspSerThrLeuGlyTyrLeuValGln	346
OY	1018	TTCGGGCGCTCCGCGGGGAGGCGCAGCGGATGAGGTGCCCGCGGCGCGCACTGC	1077
Db	347	LeuGlyProLeuGlnGlnGlySerLeuGlnHisIleValGluValProAlaGlyGlnAsnSer	366
OY	1078	ACCAAGCTGCAAGGAGCTGGCGCGCGGAGCACCGCCTACTGGTGAACGTGACCGCGCTTC	1137
Db	367	ThrThrIleGlnGlyLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe	386
OY	1138	CGCTCGGGCGCGCAGAGACCGGCTGTCCGCCAAGGCTCGACAGGCCAGCGCGCGCGCC	1197
Db	387	ArgSerGlyArgGlnArgAlaLeuSerAlaLysAlaLysThrAlaSerGlyGluArgTle	406
OY	1198	CGCGCAGCGCCGTCGCGCGCGCGCCAGACCGCGGAGACCGCCAGCGCTGAGCGCG	1251
Db	407	Arg-----ValProGlnAlaMetArgProGluAlaGlyLeuArgGluPro	421
RESULT 6			
ID	AA88340		
XX	AA88340 standard; protein; 285 AA.		
XX	AA88340;		
AC			
DT	23-MAY-2001 (first entry)		
XX			
DE	Human membrane or secretory protein clone PSEC0053.		
XX			
KM	Human; secretory protein; membrane protein; vaccine; gene therapy;		
XX	Rheumatoid arthritis; diabetes.		
OS	Homo sapiens.		
XX			
PM	EP1067182-A2.		
XX			
PD	10-JAN-2001.		

XX 07-JUL-2000; 2000BP-00114090.
 XX
 XX 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX WPI; 2001-093889/11.
 DR N-PSDB; AAF93767.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 XX gene therapy or as candidate target molecules in drug development.
 PS Claim 1; SEQ ID NO 48; 609pp + Sequence Listing; English.

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF92232 - AAF92235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can
 CC be used in vaccines. The polynucleotide sequences can be used in gene
 CC therapy. The polynucleotide sequences and the proteins they encode may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate secretory protein/membrane protein expression. The
 CC nucleic acids and complementary sequences may also be used as DNA probes
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
 CC and quantitate the presence of similar nucleic acid sequences in samples.
 CC They may also be used to study the expression and function of secretory
 CC proteins/membrane polypeptides and their role in metabolism. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC against them and in assays to identify modulators (agonists and
 CC antagonists) of expression and activity. The antibodies and antagonists
 CC may also be used as therapeutic agents to down regulate expression and
 CC activity. The antibodies may also be used as diagnostic agents for
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme
 CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
 CC treated include rheumatoid arthritis and diabetes

XX
 XX Sequence 285 AA;

Alignment Scores:
 Pred. No.: 1 94e-88 Length: 285
 Score: 1409.50 Matches: 282
 Percent Similarity: 98.9% Conservative: 0
 Best Local Similarity: 98.9% Mismatches: 2
 Query Match: 59.2% Indels: 2
 DB: 4 Gaps: 1

US-10-699-035A-5 (1-1254) x AAB88340 (1-285)

QY 1 ATGCTCCCTGGAACGGCGCTCGGCTGAGCTTGCGGCTGCGCGGAGC 60
 DB 1 MetLeuProTriPThrAlaLeuGlyLeuAlaLeuSerLeuAlaLeuAlaArgSer 20
 QY 61 GGGCGGGAAGGGGTCCACCAAGATCAGCCCCCGAGGGGAGCTGATGTTCTGCTGAGC 120
 DB 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
 QY 121 AACTGACGACGGGTCTCTACTACGAGTCTCCCGGGTTCCGGAGTTGTGGGGGACGTC 180
 DB 41 SerSerAlaSerValSerHisTyrGlyIupSerArgValArgGluPheValGlyGlnLeu 60
 QY 181 GTGGCTCACTGCGCCCTGGGCAACGGGGGCTTGCGGCAAGTCTGGTGCACGTGGGCACT 240
 DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
 QY 241 CGGCAATACCGAGATTCCCTTGGCCACAGACAGCTCGGGTGGGCTGCCAGATGCG 300

DB 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 100
 QY 301 GTGGCTGCTTGTGCCCAAGCGCATGGGTGACACCCACATGGCGCTGGCTCTATGCC 360
 DB 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
 QY 361 AAGAACAGCTGTTTGTCTGAAGCATCAGGTGCCCGGCGGAGGTCGCCAAAGTCTGGTG 420
 DB 121 LysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyValProValValLeuVal 140
 QY 421 TGGGTCAGATGGCGGCTCCAGGACCTGTGGGGCCCCCATGACAGAGCTCAAGAC 480
 DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnLeuLeuAsp 160
 QY 481 CTGGCGCTCACCGTTCATTGTGACGACCGCGGAGGCACTTCTGGAGCTGTGACCC 540
 DB 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 180
 QY 541 GCTGCTCAGCCCTGCGGAGAGACCTGCACTTTGTGACGTGATGACTGCAATC 600
 DB 181 AlaAlaSerAlaProAlaGluAlaGlySerHisPheValAspValAspAspLeuHisIle 200
 QY 601 ATTGTCCAGAGCTGAGGGGCTCCATTCTC---GGCATGGCGGCGGACAGCTCATGCC 657
 DB 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetCArgProGlnGlnLeuHisAla 220
 QY 658 ACGGATCAGCTCAGCGGCTTCCGCTGGCTGGCCACCCCTGCTGACCGGACGACTCG 717
 DB 221 ThrGluIleThrSerSerGlyPheArgGluAlaTrpProProLeuLeuThrAlaAspSer 240
 QY 718 GGCTACTATGTGCTGAGACTGTGCGCCAGCGCCGAGCGG-GGCTGCAACGCGCACA 776
 DB 241 GlyTyrTyrValLeuGlyLeuValProSerAlaGlnProGlyGlyValThrProAla 260
 QY 777 GGTGCGAGGAAACGCGACGAGCTGGATCTGGGCGGCTCGACCGGACGAGCACTAGA 836
 DB 261 AlaAlaArgGluArgGlyGlyLeuAspPheGlyArgProArgProGlyHisGlyLeuArg 280
 QY 837 CGTGCGCTAGTGCC 851
 DB 281 ArgGlyAlaSerAla 285

RESULT 7
 ADY63045
 ID ADY63045 standard; protein, 285 AA.
 XX
 XX AC ADY63045;
 XX
 XX 02-JUN-2005 (first entry)
 XX
 DE Human clone PSEC0053 protein, SEQ ID 48.
 XX
 KW Gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1514933-A1.
 XX
 PD 16-MAR-2005.
 XX
 PF 07-JUL-2000; 2004BP-00027228.
 XX
 PR 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 PR 07-JUL-2000; 2000BP-00114090.
 XX
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX WPI; 2005-203865/22.
 DX N-PSDB; ADY63044.
 DR

CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antifungal; antineumatic; antihypertoid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORX-associated disorder. The nucleic acids can be used to express ORX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypochloridism, cholesterol deficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 299 AA;

Alignment Scores:

Pred. No.: 6.32e-86 Length: 299
Score: 1373.00 Matches: 282
Percent Similarity: 94.6% Conservative: 0
Best Local Similarity: 94.6% Mismatches: 8
Query Match: 57.7% Indels: 8
DB: 3 Gaps: 3

US-10-699-035A-5 (1-1254) x AAB42581 (1-299)

QY 3 GCTCCCGGAGAGGGGCTGCGCTTGGCGCTTGGCGCTTGGCGCGGAGCGG 62
Db 9 AAlaProLeuAlaPglYAlaArg-LeuAlaLeuSerLeuAlaG-TyrArgTArg-----T 26
QY 63 CGCGAGGCGGCTCCACAGCATCAGCCCGCGAGGAGCATGATGTTCTGCTGAGCAG 122
Db 26 hrProAspCys--ProProAlaSerAlaProGlu---AspLeuMetPheLeuLeuAspSe 44
QY 123 CTGACCGCAGCGTCTCTACTACGAGTTCTCCGGGTTGGGAGTTTGTGGGCGACCTGAT 182
Db 44 rSerAlaSerValSerHisTArgLuphSerArgValArgLuphValGlyGlnLeuVal 64
QY 183 GCGTCACATGCGCCCGGAGCAGCGGGCGCTGCGTCCAGTCTGGTGGCAGTGGCAGTGG 242
Db 64 lAlaProLeuProLeuAlaPro**AlaLeuAlaArgAlaSerLeuValHisValGlySerAr 84
QY 243 GGCATACACGAGTTCCTCTGCGCAGCAGCAGTGGGCTGAGGCTGCGCAGGATGCGGT 302
Db 84 gProTyrThrArgLuphPheProPheGlyGlnHisSerSerArgLuphAlaAlaGlnAspAlaVa 104
QY 303 GCGGCTTCTGCGCAGCGGAGCGAGTGGTACACCCAGCAGTGGCGCTGGCTGTATGCCAA 362
Db 104 lArgAlaSerAlaGlnArgMetGlyAspThrHisThrArgLuphAlaLeuValTyrAlaLly 124
QY 363 GGAACAGCTGTTTGTGGAAGCATCAGTGGCGCGGCGCAGGAGGCGCCCAAGTCTGTGTG 422
Db 124 sGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuValTyr 144
QY 423 GGTGACAGATGCGCGCTCCAGCAGCAGTGGGAGCGCGCCAGCAGGAGACTAAGGAGCT 482
Db 144 pValThrAspGlyGlySerSerAspProValGlyProPrometGlnGlnLeuLysAspLe 164
QY 483 GGGCGCTACCGGTTCATTTGTGACAGCAGCGGCGGAGGCACTTCTGGAGCTGTACGCGC 542
Db 164 uGlyAlaThrValPheIleValSerThrArgLuphAspPheLeuGlnLeuSerAlaAla 184
QY 543 TGGCTCAGCGCTGCGGAGGAGCAGCAGTGGGAGCGGAGTGGAGTGCATCATCAT 602
Db 184 aAlaSerAlaProAlaGlnLuphHisLeuHisPheValaArgValaAspAspLeuHisIleL 204
QY 603 TGTCCAAAGAGTGAAGGAGCTCCATTCTC--GCGATGCGGCGGAGCAGCTCCATGCCAC 659
Db 204 eValGlnGlnLeuAlaArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAlaI 224

QY 660 GAGATACATGCTCCAGCGGCTTCCGCTGCGCTGCGCACCCTGCTGACCGAGACTCGGG 719
Db 224 rGlnLuphHisSerSerGlyPheArgLeuAlaTyrProProLeuLeuThrAlaAspSerG 244
QY 720 CTACTATGCTGAGAGCTGAGTGGCCAGCGCGGAGGCTGCAAGACGCGACGACT 779
Db 244 YTrTyrValLeuGlnLeuValProSerAlaGlnProGlyAlaAlaArgGlnGlnLe 264
QY 780 GCCAGGAGACCGCAGCAGTGGATCTGGGCGCGCTCGACCCGGAACAGCAGTACGAGCT 839
Db 284 uProLysAlaMetAlaThrAspTyrPleTyrAlaGlyLeuAspProAspThrAspTyrAspVa 284
QY 840 GCGGTATGCTGCTGAGTCCAAACGTCGCGCTCTGAGCGCCCAAGATC 885
Db 284 lAlaLeuValProGlnSerAlaMetAlaArgLeuLeuAlaArgProGlnLe 299

RESULT 9

AAB87344 ID AAB87344 standard; protein, 215 AA.

XX AAB87344;

DT 22-MAY-2001 (first entry)

XX Human gene 3 encoded secreted protein HMTB078, SEQ ID NO:85.

KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; attherosclerosis;
KW cardiovascular disorder; angioygenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerrary; cell culture;
KW chemotaxis; food additive; binding partner identification.

XX Homo sapiens.

XX WO200118022-A1.

XX 15-MAR-2001.

PF 31-AUG-2000; 2000WO-US024008.

PR 03-SEP-1999; 99US-0152315P.

PR 03-SEP-1999; 99US-0152317P.

XX (HUMA-) HUMAN GENOME SCT INC.

PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
PI Sopet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
DR WPT, 2001-203081/20.
XX N-PSDB; AAF91860.
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.

XX Claim 11; Page 532-533; 607P; English.

PS AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
XX protein genes, and AAB87342-AA87413 represent the proteins they encode.
XX AAB87414-AA87454 represent human secreted protein fragments. The genes
XX and their corresponding secreted proteins are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 52 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of

CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the proteins mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention
XX
XX

SQ Sequence 215 AA;

Alignment Scores:

Pred. No.:	5, 95e-65	Length:	215
Score:	1067.50	Matches:	213
Percent Similarity:	99.1%	Conservative:	0
Best Local Similarity:	99.1%	Mismatches:	1
Query Match:	44.9%	Indels:	1
DB:	4	Gaps:	1

US-10-699-035A-5 (1-1254) x AAB87344 (1-215)

QY 1 ATGCTCCCTGAGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCGAGC 60
Db 1 MetLeuProTrpHrAla***GlyLeuAlaLeuSerLeuAlaLeuAlaGser 20
QY 61 GCGCGCGAGCGCGCTGAGCATGAGCGCGCGAGCGAGCGAGCGAGCGAGC 120
Db 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
QY 121 AGCTGAGCGAGCGCTGAGCATGAGCGCGCGAGCGAGCGAGCGAGCGAGC 180
Db 41 SerSerAlaSerValSerHisTrpGluPheSerAlaValArgGluPheValGlyGlnLeu 60
QY 181 GTGCTCCACTGCGCGCGAGCGAGCGCGCGCGCGCGAGCGAGCGAGCGAGC 240
Db 61 ValAlaProLeuProLeuGluGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCGATACACCGAGTTCCCTTGGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 300
Db 81 ArgProTrpThrGluPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla 100
QY 301 GTCGCTGCTTTCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 360
Db 101 ValArgAlaSerValGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGAACAGCTGTTTCTGAGCATGAGCGCGCGAGCGAGCGAGCGAGCGAGCGAGC 420
Db 121 LysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyAlaProLysValLeuVal 140
QY 421 TGGGTGACAGATGGCGCTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 480
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProProMetGlnGluLeuLysAsp 160
QY 481 CTGGGCGTACCGGTGTTTCAATTGTCAGACCGCGAGCGAGCGAGCGAGCGAGCGAGC 540
Db 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 180
QY 541 GCTGCTCAGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 600
Db 181 AlaAlaSerAlaProAlaGluLysHisLeuHisPheValAspValAlaSerPheLeuHisIle 200
QY 601 ATTGTCAAAGAGCTGAGGCGCTGCATTCTC--GCGATGCGGCGG 642

Db 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgPro 215

RESULT 10.

ABG65347

ID ABG65347 standard; protein; 215 AA.

AC ABG65347;

XX 27-AUG-2002 (first entry)

DE Human albumin fusion protein #2022.

XX Albumin fusion protein; therapeutic protein X; human albumin; HA;

KW human serum albumin; HSA; cancer; reproductive disorder;

KW digestive disorder; immune disorder; endocrine disorder;

KW haematopoietic disorder; neural disorder; connective disorder;

KW cytostatic; antifertility; antiinflammatory; antitumor;

KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;

KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;

KW osteopathic; antiarthritic.

XX Homo sapiens.

OS Synthetic.

PN WO200177137-A1.

PD 18-OCT-2001.

PF 12-APR-2001; 2001WO-US011988.

PR 12-APR-2000; 2000US-0229358P.

PR 25-APR-2000; 2000US-0199384P.

PR 21-DEC-2000; 2000US-0256931P.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Haseltine WA;

PI WPI; 2002-010886/01.

DR New fusion protein for treating disease e.g. diabetes comprises an

PT albumin fused to a therapeutic protein.

XX Claim 1; Page 1935; 2102pp; English.

XX The present invention relates to albumin fusion proteins comprising a

CC therapeutic protein X and human albumin (HA, also known as human serum

CC albumin, HSA). The proteins are useful for treating a disease or disorder

CC that may be modulated by therapeutic protein X. The albumin extends the

CC shelf-life of protein X, and may increase its biological in vitro/in vivo

CC activity. The protein is useful for treating and diagnosing disorders

CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's

CC disease, ulcerative colitis), immune disorders (e.g. acquired

CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),

CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,

CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,

CC schizophrenia), and connective tissue disorders (e.g. osteoporosis, arthritis).

CC ABG65326-ABG65518 represent albumin fusion proteins of the invention

XX

SQ Sequence 215 AA;

Alignment Scores:

Pred. No.:	5, 95e-65	Length:	215
Score:	1067.50	Matches:	213
Percent Similarity:	99.1%	Conservative:	0
Best Local Similarity:	99.1%	Mismatches:	1
Query Match:	44.9%	Indels:	1
DB:	5	Gaps:	1

US-10-699-035A-5 (1-1254) x ABG65347 (1-215)

QY 1 ATGCTCCCTGAGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCGAGC 60

Db 1 MetLeuProTrpThrAla***GlyLeuValAlaSerLeuAlaGlyLeuAlaLeuAlaArgSer 20
 QY 61 GCGCGGAGCGCGGCTCCACAGCATCAGCCCGCCAGGAGGAGCTATGTTCTCGTGGAC 120
 Db 21 GlyAlaGluArgGlyProProAlaSerAlaProAlaGlyAspLeuMecPheLeuLeuasp 40
 QY 121 AGCTCAGCCAGCGTCTCTCAGTACGAGTTCCTCCGAGTTCGGAGTTCGGAGCTG 180
 Db 41 SerSerAlaSerValSerHisIerGlyIupSerArgValArgGlyIupHelaGlyIupLeu 60
 QY 181 GTGGCTCAGTCCCTGGGAGCGGAGGCGCCCTGGCGAGCTGTGACGTGGAGCT 240
 Db 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
 QY 241 CGGCGATACACCGAGTTCCTCCGAGCGAGCATCAGCTCGAGTTCGGAGTTCGG 300
 Db 81 ArgProTrpThrGluPheProPheGlyGlnHisSerSerGlyGluAlaGlnAspAla 100
 QY 301 GTGGCTGCTTCTGCGCAGCGCATYGGGTGACACCCACACTGCGCTGGCTGTATGCC 360
 Db 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
 QY 361 AAGGACAGCTGTTTGTGTCAGACATCAGTGCCTCGGCGAGGAGTTCCTCGTGGT 420
 Db 121 LysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyValProGlyValLeuVal 140
 QY 421 TGGGAGACAGATGAGGCGCTCCAGCGACCTGCGGCGCCCGCCATGACAGAGTCAAGAC 480
 Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProProMetGlnGlnLeuLysAsp 160
 QY 481 CTGGCGCTGACCGCTGTTCTATGTCAGACACCGCGCGAGGCACTTCCTGAGCTGTACGC 540
 Db 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGlnLeuSerAla 180
 QY 541 GCTGCTTCAAGCCCTGCGGAGACACCTGCACTTTGTGAGAGTGGATGACTGCACATC 600
 Db 181 AlaAlaSerAlaProAlaGlnLysHisIleHisIlePheValAspValAspAspLeuHisIle 200
 QY 601 ATTGTCCAAGAGCTGAGGAGCTCCATTCTC--GGCATCGGCGCG 642
 Db 201 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetArgPro 215
 RESULT 11
 ADL78614
 ID ADL78614 standard; protein; 215 AA.
 AC ADL78614;
 XX
 XX 20-MAY-2004 (first entry)
 DT
 DE Albumin fusion protein related therapeutic protein X, SEQ ID NO 2096.
 XX
 XX albumin fusion protein; cytostatic; antinaemic; antiarthritic;
 XX antiaesthetic; anti-HIV; immunosuppressive; antiinflammatory;
 XX antipsoriatic; antibacterial; osteoplastic; dermatological; antipout;
 XX immunomodulator; antiarrhythmic; cardiac; nootropic; antidiabetic;
 XX nephroretic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
 XX antidiabetic; anabolic; hypertensive; vulnery; gene therapy; cancer;
 XX reproductive system disorder; therapeutic protein.
 XX
 XX Undenified.
 XX OS
 XX US2004010134-A1.
 PN
 PD 15-JAN-2004.
 XX
 PF 12-APR-2001; 2001US-00833245.
 XX
 PR 12-APR-2000; 2000US-0229358P.
 PR 25-APR-2000; 2000US-0199384P.
 PR 21-DEC-2000; 2000US-0256931P.
 XX

PA (ROSE/) ROSEN C A.
 PA (HASE/) HASELTINE W A.
 XX
 PI Rosen CA, Haseeltine WA;
 XX
 DR WPI: 2004-090519/09.
 XX
 PS
 XX
 XX
 XX The invention relates to a novel albumin fusion protein. The invention
 CC further relates to: a composition comprising the albumin fusion protein
 CC and a pharmaceutical carrier; a kit comprising the composition of the
 CC albumin fusion protein formula; a method of treating a disease or
 CC disorder in a patient comprising the step of administering the albumin
 CC fusion protein; a method of treating a patient with a disease or disorder
 CC that is modulated by Therapeutic protein: X, or its fragment or variant;
 CC a method of extending the shelf life of Therapeutic protein: X, or its
 CC fragment or variant; a nucleic acid molecule comprising a polynucleotide
 CC sequence encoding the albumin fusion protein; a vector comprising the
 CC nucleic acid molecule of the albumin fusion protein; and a host cell
 CC comprising the nucleic acid molecule of the albumin fusion protein. The
 CC albumin fusion protein and its compositions have the following
 CC activities: cytostatic, antinaemic, antiarthritic, antiaesthetic, anti-
 CC HIV, immunosuppressive, antiinflammatory, antipsoriatic, antiarrhythmic,
 CC osteoplastic, dermatological, antitumor, immunomodulator, antiarrhythmic,
 CC cardiac, nootropic, antiparkinsonian, tranquilizer, antidiabetic, anabolic,
 CC neuroprotective, and vulnery. The albumin fusion protein nucleic acid may
 CC be used in gene therapy to treat disorders. The albumin fusion protein is
 CC useful for diagnosing, treating, preventing or ameliorating diseases or
 CC disorders comprising indication: Y. The diseases or disorders include:
 CC cancer (e.g. leukemia, colon, bone, breast, liver or lung cancer),
 CC immune or hematopoietic diseases (e.g. anemia, Hodgkin's disease, acute
 CC lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,
 CC autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
 CC disease), reproductive system disorders (e.g. prostatic, inguinal
 CC hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
 CC Leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,
 CC Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
 CC or cachexia), cardiovascular disease (e.g. Thrombomas, heart disease,
 CC arrhythmia, cardiac arrest, heat valve disease, hypernatraemia or
 CC hyponatraemia), mixed foetal diseases (e.g. foetal alcohol syndrome,
 CC Down's syndrome, Patou syndrome, Turner's syndrome, Apert syndrome or Tay-
 CC Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
 CC tract infections or renal disorders), neural or sensory disease (e.g.
 CC Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
 CC cerebellar ataxia, attention deficit disorder, autism or obsessive
 CC compulsive disorder), respiratory diseases (e.g. emphysema, lung cancer or
 CC occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
 CC disease or glomerulonephritis), digestive diseases (e.g. portal
 CC hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
 CC or connective tissue or epithelial diseases (e.g. Crohn's disease,
 CC scleroderma, wound healing or epidermolysis bullosa). This sequence
 CC represents a therapeutic protein X relating to the albumin fusion protein
 CC of the invention. The sequence listing data for this specification was
 CC downloaded from the USPTO website.
 CC
 XX
 XX
 SQ Sequence 215 AA;
 Alignment Scores:
 Pred. No.: 5,956-65 Length: 215
 Score: 1067.50 Matches: 213
 Percent Similarity: 99.1% Conservative: 0
 Best Local Similarity: 99.1% Mismatches: 1
 Query Match: 44.9% Indels: 1
 DB: 8 Gaps: 1
 US-10-699-035A-5 (1-1254) x ADL78614 (1-215)


```

QY 1 ATGCTCCCTGGAACGGCGCTCGGCTGAGCTTGGCGGTGGCGCGGAGC 60
DB 1 MetleuProtrpThxAla**GlyleuAlaLeuSerleuArgleuAlaLeuAlaArgSer 20
QY 61 GGCGGGAGCGGGGTCCACAGATAGCCCGCGGGGAGCTGATGTTCCGTGGAGC 120
DB 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
QY 121 AGCTGAGCAGCGCTCTCACTACGAGTTCCTCCGGGTTCCGGAGATTGTGGGGCAGCTG 180
DB 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACTGCCCCCTGGGCAACCGGGGCCCTGCGTGCAGTCTGGTGACGCTGGGCACT 240
DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCGATACACGAGTTCCTTCCCTTGGCGCAACGCTCGGGGTGGAGCTCCGAGATGGG 300
DB 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla 100
QY 301 GTGGGTCTTTCGCGCGAGCATAGGGTGACACCCACACTGCGCTGGCGGTCTATGGC 360
DB 101 ValAlaGluAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGAACAGCTGTTGCTGAAGCATGAGTCCCGCGGCGGAGGAGCCAAAGTCTGTG 420
DB 121 LysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyValProIysValLeuVal 140
QY 421 TGGGTGACAGATAGCGGCTCCGAGCCGCTGTCGGGCGGCGGCGGCGGCGGCGGCGGCG 480
DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProIleGlnLeuLysAsp 160
QY 481 CTGGGCGTCACTGTTGCTGATTCGACACCGGCGGCGGAGCACTTCTGAGCTGTAGCC 540
DB 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuLeuSerAla 180
QY 541 GTGCTCCAGCCCTTCGCGGAGAGCACTGCACTTTGTGACGCTGATGACCTGACATC 600
DB 181 AlaAlaSerAlaProAlaGluLysHisLeuHisPheValAspValAspAspLeuHisIle 200
QY 601 ATTGTCCAGAGCTGAGGGGCTTCATTC--GCCATGCGGCGG 642
DB 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgPro 215

```

RESULT 12
AAB87418 standard; protein; 242 AA.
ID AAB87418 standard; protein; 242 AA.
AC AAB87418;
XX
XX
DT 22-MAY-2001 (first entry)
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiodysplasia; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; cell culture;
KW chemotaxis; food additive; binding partner identification.
OS Homo sapiens.
PN WO200118022-A1.
XX
XX 15-MAR-2001.
XX
XX 31-AUG-2000; 2000MO-US024008.
XX

```

PR 03-SEP-1999; 99US-0152315P.
PR 03-SEP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Ni J, Baker KP, Birse CE, Piscella M, Komatsuji GA, Rosen CA;
PI Soppet DR, Young PE, Edner R, Duan DR, Olsen HS, Latleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
XX WPI; 2001-203081/20.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX Disclosure; Page 18; 607pp; English.
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87414-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiodysplasia, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX SEQ Sequence 242 AA;
XX
XX Alignment Scores:
XX Pred. No.: 6,09e-65 Length: 242
XX Score: 1067.50 Matches: 213
XX Percent Similarity: 99.1% Conservative: 0
XX Best Local Similarity: 99.1% Mismatches: 1
XX Query Match: 44.9% Indels: 1
XX DB: 4 Gaps: 1

```

US-10-699-035A-5 (1-1254) x AAB87418 (1-242)

```

QY 1 ATGCTCCCTGGAACGGCGCTCGGCTGAGCTTGGCGGTGGCGCGGAGC 60
DB 21 MetleuProtrpThxAla**GlyleuAlaLeuSerleuArgleuAlaLeuAlaArgSer 47
QY 61 GGCGGGAGCGGGGTCCACAGATAGCCCGCGGGGAGCTGATGTTCCGTGGAGC 120
DB 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 67
QY 121 AGCTGAGCAGCGCTCTCACTACGAGTTCCTCCGGGTTCCGGAGATTGTGGGGCAGCTG 180
DB 61 ValAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 67
QY 181 GTGGCTCCACTGCCCCCTGGGCAACCGGGGCCCTGCGTGCAGTCTGGTGACGCTGGGCACT 240

```

Db 88 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 107
Qy 241 CGGCCATACACCGAGTTCCCTTGGCCAGCAGCTCGGGTGAAGCTGCCAGGATGCG 300
Db 108 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaGlnAspAla 127
Qy 301 GTGGCTGCTTTGGCCAGCGCATGGGTGACACCCACACTGGCTGGCTGTATGCC 360
Db 128 ValArgAlaSerAlaGlnArgMetClyAspThrHisThrGlyLeuAlaLeuValTyrAla 147
Qy 361 AAGGAACAGCTGTTTGTGAAGCATCAGTGCCTCCGAGGGGTGCCAAAGTGTGGTG 420
Db 148 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProlyValLeuVal 167
Qy 421 TGGGTGACAGATGGCGGCTCCAGCGACCTTGGGCCCCCATGACAGAGCTCAAGAC 480
Db 168 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGluLeuAsp 187
Qy 481 CTGGGCGTCACCGCTTATTGTTCAGCACCGGCGAGGCACTTCTGGAGCTGTCAACC 540
Db 188 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 207
Qy 541 GCTGCTCAGGCCCCCGAGAGACCTGCACCTTGTGAGCGTGGATGACCTGCACATC 600
Db 208 AlaAlaSerAlaProAlaGlnIuYshIleuHisPheValAspValAspAspLeuHisIle 227
Qy 601 ATTGTCAAGAGCTGAGGGGCTCCATTCTC--GCGATGCGGCGC 642
Db 228 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgPro 242

RESULT 13
AAB87424
ID AAB87424 standard; protein; 226 AA.
XX
AC AAB87424;
XX
DT 22-MAY-2001 (first entry)
DE
XX Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;
KW inflammatory; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vunerary; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
XX Homo sapiens.
XX
XX WO200118022-A1.
XX
XX 15-MAR-2001.
XX
XX 31-AUG-2000; 2000WO-US024008.
XX
XX 03-SEP-1999; 99US-015231SP.
XX
XX 03-SEP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA,
XX Sopet DR, Young PE, Edner R, Duan DR, Olsen HS, Lafleur DW,
XX Moore PA, Shi Y, Wei Y, Florence KA;
XX
XX WPI, 2001-203081/20.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
XX Parkinson's diseases and cancers.

XX
PS Disclosure; Page 18; 607pp; English.
XX
XX AAF1558-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87332-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
CC
XX
SQ Sequence 226 AA;
Alignment Scores:
Pred. No.: 1.44e-64 Length: 226
Score: 1062.00 Matches: 210
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 44.6% Indels: 0
DB: 4 Gaps: 0
US-10-699-035A-5 (1-1254) x AAB87424 (1-226)
Qy 1 ATGTCCCTCGAAGCGGCGCTGCGCTGAGACTTGGCGAGTTCGCGCGAGC 60
Db 14 MetLeuProTyrThrAlaLeuGlyLeuAlaLeuSerLeuAlaLeuAlaArgSer 33
Qy 61 GACGCGAGCGCGGTCCACACGATCAGCCCCCGAGGGACCTGATGTTCTGTGAC 120
Db 34 GlnAlaGluArgGlyProProAlaSerAlaProAlaArgIlyAspLeuMetPheLeuLeuAsp 53
Qy 121 AGCTCAGCAGCGCTGTCATAGAGTTCCTCCGGGTTCGGGAGTTTGGGGAGCGTG 180
Db 54 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 73
Qy 181 GTGGCTCAGTGCCTTGGGACACCGGGGCGCTGCTGCAGTCTGGTGCACAGT 240
Db 74 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 93
Qy 241 CGGCCATACACCGAGTTCCCTTGGCCAGCAGCTCGGGTGAAGCTGCCAGGATGCG 300
Db 94 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaGlnAspAla 113
Qy 301 GTGGCTGCTTTGGCCAGCGCATGGGTGACACCCACACTGGCTGGCTGTATGCC 360
Db 114 ValArgAlaSerAlaGlnArgMetClyAspThrHisThrGlyLeuAlaLeuValTyrAla 133
Qy 361 AAGGAACAGCTGTTTGTGAAGCATCAGTGCCTCCGAGGGGTGCCAAAGTGTGGTG 420
Db 134 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProlyValLeuVal 153

QY 421 TGGGTGACGATGCGCGCTCCAGCCAGCTGTGGGCCCCCGCATGACGAGCTCAAGAGC 480
DB 154 TTPVATHTHASPGLYGLYSERSEAPROVALGILYPROBOWEGLINLEULYASAP 173
QY 481 CTGGGGCTACCGGTTCATTGTGACGACCGCGGACACTCTCTGAGCTGTACGCC 540
DB 174 LGGVYVATHTHVALPHEILEVALSERTHGTYAGGLYASPHLEULLEUSERALA 193
QY 541 GGTGCTCAGCCCTGCGGACGACGACCTGACCTTTGTGACGCTGATGACCTGCATC 600
DB 194 AAlaIaSerAlProAlaGluLulYSHISLeuHisPheValAspValAspAspLeuHisIle 213
QY 601 ATTGTCAAGAGCTGAGGGGCTTCATTCTC 630
DB 214 ILeValGInGluLeuArgILYSerIleLeu 223
RESULT 14
AAB87419
ID AAB87419 standard; protein, 186 AA.
AC AAB87419;
XX
XX
DT 22-MAY-2001 (first entry)
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:160.
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery; cell culture;
KW chemocaxis; food additive; binding partner identification.
XX
XX
OS Homo sapiens.
XX
XX
PN MO200118022-A1.
XX
XX
PD 15-MAR-2001.
XX
XX
PF 31-AUG-2000; 2000WO-US024008.
XX
XX
PR 03-SEP-1999; 99US-015231SP.
XX
XX
PR 03-SEP-1999; 99US-0152317P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
PI Sopet DR, Young PE, Ebnert R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA.
XX
XX
DR WPI; 2001-203081/20.
XX
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX
PS Disclosure; Page 18; 607P; English.
XX
XX
CC AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87419-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental

CC abnormalities, hematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, kidney disorders,
CC cardiovascular disorders, angiogenic disorders, endocrine
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX
SQ Sequence 186 AA;
XX
XX
Alignment Scores:
Pred. No.: 5 33e-54 Length: 186
Score: 908.00 Matches: 156
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 38.2% Indels: 0
DB: 4 Gaps: 0
US-10-699-035a-5 (1-1254) x AAB87419 (1-186)
QY 472 GCTCTGACATGAGGGGGGCGCCACAGGAGCTGTGACCGGCATCTGTACACACAGCA 413
DB 1 AAlaProAlaTrpGlyGlyProGlnGlyArgTrpSerArgHisLeuSerProThrProAla 20
QY 412 CTTTGGGACCCCTGCGGGGACCTGATGCTTCAGCAAAACGCTTCTTGGCATTA 353
DB 21 LeuTrpAlaProLeuAlaGlyHisLeuMetLeuGlnGlnThrAlaValProTrpHisArg 40
QY 352 CGAGGCGACAGGCGACATGAGGGGTGTGACCCATGCGGTGGGACAGACGACGACCTCT 293
DB 41 ProAlaProGlyGlnCysGlyHisProCysValaGlyGlnHisAlaProHisPro 60
QY 292 GGGCAGCTCACCAGCTGTGCTGCGGACGAGGAGAACTCGGTGTATGCGGACCTGCC 233
DB 61 GlyGlnProHisProSerCysHisAlaGlyArgGlyThrArgCysMetAlaAspCysPro 80
QY 232 CGTGACACAGACTGACGACGACGCGCGCTGCCAGGGGACGTGAGACCCACAGCTGCC 173
DB 81 ArgAlaProAspTrpHisAlaGlyProArgCysProGlyAlaValaGlnProProAlaIla 100
QY 172 CCACAAACTCCCGAACCAGGAGAACTGTGATGAGAGACGCTGCTGAGCTGTCCAGCA 113
DB 101 ProGlnThrProGlnProGlyArgThrArgSerGluArgArgTrpLeuSerCysProAla 120
QY 112 GGAACATCAGTCCCTCGGGGGCTGATGTGCTGAGACCGGCGCTCCGCGCTCCGCG 53
DB 121 GlyThrSerGlyProLeuGlyGlyLeuMetLeuValaAspArgAlaProArgArgSerAla 140
QY 52 CCAGGCGACCGCGACAGCTCAGGGGCGAGGCGAGGCGCGCTCCAGGGAGCA 2
DB 141 ProAlaProAlaIaSerSerGlyProGlyArg***ProSerArgGlyAla 157
RESULT 15
AAB32500
ID AAB32500 standard; protein, 180 AA.
AC AAB32500;
XX
XX
DT 24-MAR-2003 (first entry)
XX
XX
DE Human von Willebrand Factor A (VA) domain.
XX
XX

KW Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
 KW extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
 KW gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 1..2
 FT /note= "Encoded by GGC"
 FT Misc-difference 179..180
 FT /note= "Encoded by CTC"
 XX
 XX WO20028184-A1.
 XX
 XX PD 07-NOV-2002.
 XX
 XX PF 02-MAY-2002; 2002WO-AU000542.
 XX
 XX PR 02-MAY-2001; 2001AU-00004701.
 XX
 XX PA (MURDOCH CHILDRENS RES INST.
 XX
 XX PI Bateman JF, Fitzgerald DJ;
 XX
 XX DR WPI; 2003-111873/10.
 XX
 XX DR N-PSDB; AAD50397.
 XX
 XX PT New isolated Willebrand Factor A-Related Protein polypeptide useful for
 PT the manufacture of a medicament in the treatment of a disease condition
 PT of the extracellular matrix, in particular arthritis.
 XX
 XX PS Claim 7, Page 72-73; 103pp; English.
 XX
 XX CC The invention relates to Willebrand Factor A domain related-protein
 CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
 CC superfamily of extracellular matrix (ECM) proteins. WARP is used as an
 CC molecular marker, used for detecting a loss of ECM integrity in an animal
 CC subject, monitoring repair, regeneration or other disease processes in an
 CC animal subject and detecting a disease condition or a propensity for the
 CC development of a disease condition in an animal subject. The invention is
 CC useful for the manufacture of a medicament in the treatment of a disease
 CC condition of the ECM. The disease condition involves the cartilage, and
 CC is preferably arthritis. The invention is also used in gene therapy. The
 CC present sequence is human VA domain
 XX
 XX SQ Sequence 180 AA;
 Alignment Scores:
 Pred. No.: 6,21e-54 Length: 180
 Score: 907.00 Matches: 179
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 38.1% Indels: 0
 DB: 6 Gaps: 0
 US-10-699-035A-5 (1-1254) x AAE32500 (1-180)
 QY 94 CGAGGGGACCTGATGTTCTGCTGACAGCTCAGCCAGCGTCTCTACATGAGATTCTCC 153
 Db 1 ArgGlyAspLeuWetcheLeuWaspSerAlaSerValSerHisTyrGluPheSer 20
 QY 154 CGGGTTCCGGAGATTGTTGGGGCAGCTGTGCTCCACTGCCCCCTGGGCAACGGGGCCCTG 213
 Db 21 ArgValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeu 40
 QY 214 CGTGCCAGTCTGGTGCACGTGGGCACTGCGCCATACACGAGTCCCTTCGGCCAGCAC 273
 Db 41 ArgAlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHis 60
 QY 274 AGCTCGGAGGAGCTGCCAGGATGGCGGTGGCTTGTGCCAGGCGCATGGGTCACACC 333
 Db 61 SerSerGlyGlnAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThr 80

QY 334 CACACTGGCTGGCGGCTGGTCTATGCCAAGAACAGCTGTTTCTGTAAGCATCAGGTGCC 393
 Db 81 HisThrGlyLeuAlaLeuValTyrAlaGlyGlnLeuPheAlaGluAlaSerGlyAla 100
 QY 394 CGGCCAGGGGTGCCCAAGTGTGTTGGGTGACAGATGGCGGCTCCAGGACCCCTGTG 453
 Db 101 ArgProGlyValProLysValLeuValTyrValThrAspGlyGlySerSerAspProVal 120
 QY 454 GGGCCCCCATGCAAGAGCTCAAGGACCTGGGCGTCAACCGTTCATTGTACAGCACCGGC 513
 Db 121 GlyProMetGlnGlnLeuLysAspLeuGlyValThrValPheIleValSerThrGly 140
 QY 514 CGAGCAACTTCTCTGAGCTGTCAAGCGCTGCTGACCGCTCCAGCCCTCCGAGAACAGCCTGCAC 573
 Db 141 ArgGlyAsnPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHis 160
 QY 574 TTTGTGACGTGGATGACTGTCACATCAATTGTCCAGAGCTGAGGGGCTCCATTCTC 630
 Db 161 PheValAspValAspAspLeuHisIleIleValGlnGlnLeuArgGlySerIleLeu 179

Search completed: February 13, 2006, 13:23:41
 Job time : 228.052 secs

Db 1 MetLeuProTPrThAlaLeuGlyLeuAlaLeuSerLeuArgLeuAlaArgSer 20
Qy 61 GGGCGGAGCGCGCTCCACACAGATCAGCCCCGAGGGAGCTGATGTTCTGCTGAGAC 120
Db 21 G1yAlaG1uArgG1yProProAlaSerAlaProArgG1yAspLeuMetPheLeuLeuAsp 40
Qy 121 AGCTCAGCGAGCGCTCTCCTACCTACGAGATTCTCCCGGATTCCGGAGTTTGCGGGACGCTG 180
Db 41 SerSerAlaSerValSerIleTyrG1uPheSerArgValArgG1uPheValG1yG1uLeu 60
Qy 181 GTGGCTCAGCTGCCCCCTGGGGCAGCGGGGCGCTGGCGAGCTGAGTGCATGCGGAGT 240
Db 61 ValAlaProLeuProLeuG1yThrG1yAlaLeuArgAlaSerLeuValHisValG1ySer 80
Qy 241 CGGCGCATACCCAGATTCCCTTCGCGCCAGCAGCTCGGGTGAAGCTCCCGAGGATGCG 300
Db 81 ArgProTyrThrG1uPheProPheG1yG1uHisSerSerG1yG1uAlaAlaG1uAspAla 100
Qy 301 GTGGCTGCTTTCGCCAGCGGCACTGGGTGACACCCACACTGCGCTGGCGCTGCTATGCG 360
Db 101 ValArgAlaSerAlaG1uArgMetG1yAspThrHisThrG1yLeuAlaLeuValTyrAla 120
Qy 361 AAGGAACAGCTGTTGCTGAAGCATCAGGTGCGCGGCGAGGGGTCGCCAAAGTGGTGG 420
Db 121 LysG1uG1uLeuPheAlaG1uAlaSerG1yAlaArgProG1yValProLysValLeuVal 140
Qy 421 TGGGTGACAGATGCGCGCTCCAGCGACCTCTGCGGCGCCCGCATCGAGGCTCAAGAC 480
Db 141 TrpValThrAspG1yG1ySerSerAspProValG1yProProMetG1uG1uLeuLysAsp 160
Qy 481 CTGGGCGCTCACCGTGTTCATTGTTCAGCACCGGCGCGAGGCACTTCTGGAGCTGTACGCC 540
Db 161 LeuG1yValThrValPheIleValSerThrG1yArgG1yAspPheLeuG1uLeuSerAla 180
Qy 541 GCTGCTCAGCCCGCGCGAGAAAGCACTGGAAGCTGAGAGTGCATGACACATC 600
Db 181 AlaAlaSerAlaProAlaG1uLysHisLeuHisPheValAspValAspAspLeuHisIle 200
Qy 601 ATTGTCCAAGACTGAGGGGCTCCATTCTC--GCGATGCGGCGCGAGCGCTCCATGCC 657
Db 201 IleValG1uG1uLeuArgG1ySerIleLeuAspAlaMetArgProG1uG1uLeuHisAla 220
Qy 658 ACGGAGATCAGCTCCAGCGGCTTTCGCGCTGCGCTGCGCCACTTGTGTGACCGCAAGCTCG 717
Db 221 ThrG1uIleThrSerSerG1yPheArgLeuAlaTrpProProLeuLeuThrAlaAspSer 240
Qy 718 GCGTCACTATGCTGAGAGCTGCTGCCAGCGCGCGCGCGGCGCTGCAAGAGCGACGAG 777
Db 241 G1yTyrTyrValLeuG1uLeuValProSerAlaG1uProG1yAlaAlaArgArgG1uG1u 260
Qy 778 CTGCCAGGGAACGCCACGCACTGATCTTGGCGCGGCTGAGCCCGGACACGCACTACGAC 837
Db 261 LeuProG1yAspAlaThrAspTrpIleTrpAlaG1yLeuAspProAspThrAspTyrAsp 280
Qy 838 GTGGCGCTAGTGCCTGAGTCCAAAGTGGCGCTCTGAGGCGCCAGATCTCTGGGGTGGC 897
Db 281 ValAlaLeuValProG1uSerAsnValArgLeuLeuArgProG1uIleLeuArgValArg 300
Qy 898 ACGCGCGCAGAGAGCGCGGCGCGAGCGCATGTCATCTCCACGCGCGCGCGCGCGAC 957
Db 301 ThrArgProG1uG1uAlaG1yProG1uArgIleValIleSerHisAlaArgProArgSer 320
Qy 958 CTCCCGTAGTGGGCGCCAGCGCTGAGCGCTGAGCGCGGCGCTCGGCTACACAGCGAG 1017
Db 321 LeuArgValSerTrpAlaProAlaLeuG1ySerAlaAlaAlaLeuG1yTyrHisValG1u 340
Qy 1018 TTCCGGCGCGCTGCGGGGCGGGAGCGCAGCGGCTGAGAGTGCCTCCGGGCGCGCACTGC 1077
Db 341 PheG1yProLeuArgG1yG1yG1uAlaG1uArgValG1uValProAlaG1yArgAsnG1y 360
Qy 1078 ACCAGCGTGCAGAGGCTGCGCGCGCGGACACCGCTACCTGAGTGCATGACCGCGCTTC 1137
Db 361 ThrThrLeuG1uG1yLeuAlaProG1yThrAlaTyrLeuValThrValThrAlaAlaPhe 380

Qy 1138 CGCTCGGCGCGCGAGAGCGCGCTGTCCGCAAGGCTGCAAGCCCGAGCGCGCGCGCG 1197
Db 381 ArgSerG1yArgG1uSerAlaLeuSerAlaLysAlaG1ySerProAspG1yProArgPro 400
Qy 1198 CGCCCAAGCG 1251
Db 401 ArgProArgProValProArgAlaProThrProG1yThrAlaSerArgG1uPro 418
RESULT 2
US-10-699-035A-20
; Sequence 20, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-20
Alignment Scores:
Pred. No.: 1,33e-115 Length: 418
Score: 2137.50 Matches: 417
Percent Similarity: 99.84 Mismatches: 0
Best Local Similarity: 99.88 Mismatches: 0
Query Match: 89.84 Indels: 1
Gaps: 1
US-10-699-035A-5 (1-1254) X US-10-699-035A-20 (1-418)
Qy 1 ATGCTCCCTGAGACGCGGCTGAGGCTGGCGCTGAGCTTGGCGCTGAGCGCGAGC 60
Db 1 MetLeuProTPrThAlaLeuG1yLeuAlaLeuSerLeuArgLeuAlaArgSer 20
Qy 61 GCGCGGAGCGCGCTCCACACAGATCAGCCCCGAGGGAGCTGATGTTCTGCTGAGAC 120
Db 21 G1yAlaG1uArgG1yProProAlaSerAlaProArgG1yAspLeuMetPheLeuLeuAsp 40
Qy 121 AGCTCAGCGAGCGCTCTCCTACCTACGAGATTCTCCCGGATTTCGGAGTTTGCGGGACGCTG 180
Db 41 SerSerAlaSerValSerIleTyrG1uPheSerArgValArgG1uPheValG1yG1uLeu 60
Qy 181 GTGGCTCAGCTGCCCCCTGGGGCAGCGGGGCGCTGGCGAGCTGAGTGCATGCGGAGT 240
Db 61 ValAlaProLeuProLeuG1yThrG1yAlaLeuArgAlaSerLeuValHisValG1ySer 80
Qy 241 CGGCGCATACCCAGATTCCCTTCGCGCCAGCAGCTCGGGTGAAGCTCCCGAGGATGCG 300
Db 81 ArgProTyrThrG1uPheProPheG1yG1uHisSerSerG1yG1uAlaAlaG1uAspAla 100
Qy 301 GTGGCTGCTTTCGCCAGCGGCACTGGGTGACACCCACACTGCGCTGGCGCTGCTATGCG 360
Db 101 ValArgAlaSerAlaG1uArgMetG1yAspThrHisThrG1yLeuAlaLeuValTyrAla 120
Qy 361 AAGGAACAGCTGTTGCTGAAGCATCAGGTGCGCGGCGAGGGGTCGCCAAAGTGGTGG 420
Db 121 LysG1uG1uLeuPheAlaG1uAlaSerG1yAlaArgProG1yValProLysValLeuVal 140
Qy 421 TGGGTGACAGATGCGCGCTCCAGCGACCTCTGCGGCGCCCGCATCGAGGCTCAAGAC 480
Db 141 TrpValThrAspG1yG1ySerSerAspProValG1yProProMetG1uG1uLeuLysAsp 160


```

QY      481 CTGGGCGTCAACCGTGTTCATGTTCAGACACCGCCGAGGCACTTCTGAGCTGTACGCC 540
      161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 180
QY      541 GCTGCTCAGCCCTCCGAGGAGACCTTCACCTTTGTGACCTGGATGACCTGCACATC 600
      181 AlaAlaSerAlaProAlaGluValSerHisLeuHisPheValAspValAspPheLeuHisIle 200
QY      601 ATTGTCGAAGAGCTGAGGGGCTCCATTCTC---GCGATGCGCGCGAGCAGCTCCATGCC 657
      201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
QY      658 ACGAGATCAAGCTCAGCGGCTTCCGCTGCGCTGCGACCCCTGTGACCGGACATCG 717
      221 ThrGluIleThrSerSerGlyPheArgLeuValIleTyrProProLeuLeuThrAlaAspSer 240
QY      718 GGTACTATGTGTGTGAGCTGTGTGCTGCGCCAGCGCCGAGCGGGGCTGCAAGACCCAGCAG 777
      241 GlyTyrTyrValLeuGluLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
QY      778 CTGCGCAGGGAACGCGACGCGATCTGGGCGCGGCTGACCGCGGACAGACTACGAC 837
      261 LeuProGlyAsnAlaThrAspTrpIleTyrAlaGlyLeuAspProAspThrAspTyrAsp 280
QY      838 GTGGCGCTAGTGTGCTGAGTCCACGATCGCGCTCTGAGCGCCGACATCTCGCGGTGCGC 897
      281 ValAlaIleValAlaProGluSerAsnValArgLeuLeuArgProGlnIleLeuArgValArg 300
QY      898 ACGCGCGCAGAGAGAGCGCGGCGCGACAGCGCATCTGATCTCCACGCGCGCGCGCGCAGC 957
      301 ThrArgProGluGluAlaGlyProGluArgIleValIleSerHisAlaArgProArgSer 320
QY      958 CTCGCGGTGAGTGGGCGCGCGCGCGCTGAGCGCTGACCGCGCGCGCTCGGCTACGACGTGAC 1017
      321 LeuArgValSerTrpAlaProAlaLeuGlySerAlaAlaLeuGlyTyrHisValGln 340
QY      1018 TTGCGGCGCGTGGCGCGCGGAGCGCGACGCGGTGAGGTGCGCGCGCGCGCGCAATCGC 1077
      341 PheGlyProLeuArgGlyGlyGluAlaGlnArgValGluValProAlaGlyArgAsnGly 360
QY      1078 ACCAGCTGACAGGCGCTGCGCGCGCGCGACCGCTTACTGCTGACGTCGACCGCGCTTC 1137
      361 ThrThrLeuGlnGlnIleuAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPhe 380
QY      1138 CGTGGCGCGCGAGAGCGCGCTGCGCGCGCAAGCGCTGACGCGCGCGCGCGCGCGCG 1197
      381 ArgSerGlyArgGluSerAlaLeuSerAlaValAlaCysThrProAspGlyProArgPro 400
QY      1198 CGCCACAGCCCGCTGCGCGCGCGCGCGCGAGCGCGGAGACCGCGCGCTGAGCGCG 1251
      401 ArgProArgProValProAlaArgAlaProThrProGlyThrAlaSerArgGluPro 418
      Db
      RESULT 3
      US-10-699-035A-4
      ; Sequence 4, Application US/10699035A
      ; Publication No. US20040214349A1
      ; GENERAL INFORMATION:
      ; APPLICANT: Bateman, John
      ; APPLICANT: Fitzgerald, David
      ; TITLE OF INVENTION: A Molecular Marker
      ; FILE REFERENCE: A36056 PCT USA A 071838.0142
      ; CURRENT APPLICATION NUMBER: US/10/699,035A
      ; CURRENT FILING DATE: 2003-10-31
      ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
      ; PRIOR FILING DATE: 2002-05-02
      ; PRIOR APPLICATION NUMBER: AU PR4701/01
      ; NUMBER OF SEQ ID NOS: 40
      ; SOFTWARE: FastSeq for Windows Version 4.0
      ; SEQ ID NO 4
      ; LENGTH: 415
      ; TYPE: PRT

```

```

; ORGANISM: Mus musculus
US-10-699-035A-4
Alignment Scores:
Pred. No.: 7,42e-87
Score: 1640.50
Percent Similarity: 85.9%
Best Local Similarity: 77.7%
Query Match: 68.9%
DB: 4 Gaps: 1
US-10-699-035A-5 (1-1254) x US-10-699-035A-4 (1-415)
QY      1 ATGCTCCCTGAGCGGCGCTCGGCTGAGCTTCGCGCTGCGCGGAGC 60
      1 MetLeuPheTrpThrAlaPheSerMetAlaLeuSerLeuArgGluAlaValArgSer 20
      Db
QY      61 GCGCGGAGACGGGCTCACCAGCATTAAGCCCCCGAGGGGACCTGATGTTCTGCTGAGC 120
      21 SerIleGluArgGlySerThrAlaSerAspProGlnGlyAspLeuLeuPheLeuLeuAsp 40
      Db
QY      121 AGCTCAGCAGCGTCTCTCACTACGAGTTCCTCCGGGTTCCGGAGATTGTGGGGCAGCTG 180
      41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
      Db
QY      181 GTGGCTCACTGCCCTCGGCGACCGGGGCGCTGCGTGCAGTCTGTGCACTGTGGCAGT 240
      61 ValAlaIleThrMetSerPheGlyProGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
      Db
QY      241 CCGCCATCAACGCGAGTTCCTCCCTTGGCGCGACAGCAGTCTGGGTGAGGCTCCGAGATGC 300
      81 GlnProHisThrGluPheThrPheAspGlnTyrSerSerGlyGlnAlaIleArgAspAla 100
      Db
QY      301 GTGCGGTCTTGTGCGCGCGCATGAGGTGACACCGACACATGCTGCGCTGCTGTATGCC 360
      101 IleArgValAlaProGlnArgMetGlyAspThrAsnThrGlyLeuAlaLeuAlaTyrAla 120
      Db
QY      361 AAGGAACAGCTGTTTGTGCAAGCATCAGGTGCCCGGCGAGGGGTGCCAAAGTGTCTGTG 420
      121 LysGluGlnLeuPheAlaGluGluAlaGlyAlaArgProGlyValProLysValLeuVal 140
      Db
QY      421 TGGGTACAGATGAGCGGCTTCCAGCGACCTGTGAGGCGCCCGCATGACGAGACCTCAAGAC 480
      141 TrpValThrAspGlyGlySerSerAspProValGlyProMetGlnGluLeuLysAsp 160
      Db
QY      481 CTGGGCGTCAACCGTGTTCATGTTCAGACACCGCGCGGCAACTTCTGAGCTGTACGCC 540
      161 LeuGlyValThrIlePheIleValSerThrGlyArgGlyAsnLeuLeuGluLeuAla 180
      Db
QY      541 GCTGCTCAGCCCTCCGAGGAGACCTTCACCTTTGTGACCTGGATGACCTGCACATC 600
      181 AlaAlaSerAlaProAlaGluValSerHisLeuHisPheValAspValAspPheLeuHisIle 200
      Db
QY      601 ATTGTCGAAGAGCTGAGGGGCTCCATTCTC---GCGATGCGCGCGAGCAGCTCCATGCC 657
      201 IleAlaArgGluLeuArgGlySerIleThrAspAlaMetGlnProGlnGlnLeuHisAla 220
      Db
QY      658 ACGAGATCAAGCTCAGCGGCTTCCGCTGCGCTGCGACCCCTGTGACCGGACATCG 717
      221 SerGluValLeuSerSerGlyPheArgLeuSerTrpProProLeuLeuThrAlaAspSer 240
      Db
QY      718 GGTACTATGTGTGTGAGCTGTGTGCTGCGCCAGCGCCGAGCGGGGCTGCAAGACCCAGCAG 777
      241 GlyTyrTyrValLeuGluLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
      Db
QY      778 CTGCGCAGGGAACGCGACGCGATCTGGGCGCGGCTGACCGCGGACAGACTACGAC 837
      261 LeuProGlyAsnAlaThrAspTrpIleTyrAlaGlyLeuAspProAspThrAspTyrAsp 280
      Db
QY      838 GTGGCGCTAGTGTGCTGAGTCCACGATCGCGCTCTGAGCGCCGACATCTCGCGGTGCGC 897
      281 ValSerLeuLeuProGluSerAsnValHisLeuLeuArgProGlnHisValArgValArg 300
      Db

```



```
OY 898 ACGCGCCAGAGAGGCGCGGCGCAGAGCGCATGTCATCTCCAGCGCCGCGCGCGAGC 957
|||
Db 301 ThrleuGlnGlnGlnAlaGlyProGlnuArgIleValIleSerHisAlaArgProArgSer 320
OY 958 CTCGCGGTGAGTTGGGAGCCCGCAGCGCTGAGCGCGCGCGCTCGGCTACACAGCTGACG 1017
|||
Db 321 LeuArgValSerTTrpAlaProAlaLeuGlyProArgSerAlaLeuGlyTyrHisValGln 340
OY 1018 TTCGGGCGCGCTGCGGGGCGGGAGCGCAGCGGCTGAGTGCCTCCGCGGCGCGCAATGC 1077
|||
Db 341 LeuGlyProLeuGlnGlySerLeuGlnuArgValGlnuValProAlaGlyGlnuSer 360
OY 1078 ACCAGCGTGCAGAGGCGCTGCGCGCGCGCGCAGCGCTGACCTGACCGCGCTTC 1137
|||
Db 361 ThrThrValGlnGlyLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe 380
OY 1138 CGCTGCGGCGCGGAGAGCGCGCTGTCGCGCAGAGCGCTGACCGCGCGCGCGCGCG 1197
|||
Db 381 ArgSerGlyArgGlnArgAlaLeuSerAlaValAlaCysThrAlaSerGlyAlaArgThr 400
OY 1198 CGCGCCAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1244
|||
Db 401 Arg--AlaProGlnSerMetArgProGlnuAlaGlyProArgGlnuPro 415
```

RESULT 4

```
US-10-699-035A-21
; Sequence 21, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT US/A 071838, 0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU FR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-21
```

Alignment Scores:

```
Pred. No.: 7,42e-87 Length: 415
Score: 1640.50 Matches: 324
Percent Similarity: 85.9% Conservative: 34
Best Local Similarity: 77.7% Mismatches: 56
Query Match: 68.9% Indels: 3
DB: 4 Gaps: 1
```

US-10-699-035A-5 (1-1254) x US-10-699-035A-21 (1-415)

```
OY 1 ATGCTCCCTGAGAGCGGCTCGGCTGAGCTTGCAGCTGCGCTGCGCGCGAGC 60
|||
Db 1 MetLeuPheTrpThrAlaPheSerMetAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
OY 61 GCGCGGAGCGCGCTGCACACAGCATGACCCCCGAGGGGAGCTGATGTTCTCTGTGAC 120
|||
Db 21 SerIleGluArgGlySerThrAlaSerAspProGlnGlyAspLeuLeuPheLeuLeuAsp 40
OY 121 AGCTCAGCGAGGCTCTCTCACTACAGATTCTCCCGGTTCCGGAGTTTGTGGGCGAGCTG 180
|||
Db 41 SerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlnuPheValGlyGlnu 60
OY 181 GTGGCTCACTGCGCCCTGCGGCGCGCGCGCGCTGCGCGAGCTGAGTGCAGCTGGGCGAGT 240
|||
Db 61 ValAlaThrMetSerPheGlyProGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
```

```
OY 241 CGGCGATACACCGAGTTCCCTTGCGCCAGCAGACAGCTGGGTGAGGCTGCCAGAGATGCG 300
|||
Db 81 GlnProHisThrGlnPheThrPheAspGlnTyrSerSerGlyGlnAlaIleArgPheAla 100
OY 301 GNGCGTGTCTTGCCCGCAGCGCATGGGTGACACCCACATGCGCTGCGCGCTGATGATGCC 360
|||
Db 101 IleArgValAlaIaProGlnArgMetClyAspThrAsnThrGlyLeuAlaIleuAlaIleVal 120
OY 361 AAGGAACAGCTGTTTGTGAAGCATCAGTGCCTCCGCGGAGGGGTGCCAAAGTGTGTTG 420
|||
Db 121 LysGlnGlnuLeuPheAlaGlnGlnuAlaGlyAlaArgProGlyValProIleValLeuVal 140
OY 421 TGGGTGACAGATGGGCGCTCCAGCGACCTGTGGGCGCGCGCGCGCGCGCGCGCGCGCG 480
|||
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGlnuLeuLysAsp 160
OY 481 CTGGCGCTGACCGGTTCATGTCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
|||
Db 161 LeuGlyValThrIlePheIleValSerThrGlyArgIleuLeuLeuGlnuLeuAla 180
OY 541 GCTGCGCTGAGCGCGCTGCGAGAGACCTGCACTTGTGACGTGATGACCTGCACATC 600
|||
Db 181 AlaAlaSerAlaProAlaGlnuLysHisLeuHisPheValAspValAspAspLeuProIle 200
OY 601 ATTGTCAGAGCTGAGGGGCTCCATTTCTC--GGGATGCGGCGCGCAGAGCTCCATGCC 657
|||
Db 201 IleAlaArgGlnuLeuArgGlySerIleThrAspAlaMetGlnProGlnGlnuLeuHisAla 220
OY 658 ACGAGATCAGGTCCAGCGGCTTCCGCGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
|||
Db 221 SerGlnuValLeuSerSerGlyPheArgLeuSerTrpProProLeuLeuThrAlaAspSer 240
OY 718 GGTACTATGTGCTGAGACTGGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
|||
Db 241 GlyTyrTyrValIleuGlnuLeuValProSerGlyLysAlaIleThrThrArgArgGlnGln 260
OY 778 CTGCCAGGGAACGCCAGCATGATCTGGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 837
|||
Db 261 LeuProGlyAsnAlaThrSerTrpThrThrAspLeuAspProAspThrAspTyrGln 280
OY 838 GTGGCGTACGAGCTGAGTCCAAAGTGCAGCTGCGCGCTGAGAGCGCGCGCGCGCGCGCG 897
|||
Db 281 ValSerLeuLeuProGlnSerHisValHisLeuLeuArgProGlnHisValArgValArg 300
OY 898 ACGCGCCAGAGAGAGCGCGGCGCAGAGCGCATGTCATCTCCAGCGCGCGCGCGCGAGC 957
|||
Db 301 ThrleuGlnGlnGlnAlaGlyProGlnuArgIleValIleSerHisAlaArgProArgSer 320
OY 958 CTCGCGGTGAGTTGGGAGCCCGCAGCGCTGAGCGCGCGCGCTCGGCTACACAGCTGACG 1017
|||
Db 321 LeuArgValSerTTrpAlaProAlaLeuGlyProArgSerAlaLeuGlyTyrHisValGln 340
OY 1018 TTCGGGCGCGCTGCGGGGCGGGAGCGCAGCGGCTGAGTGCCTCCGCGGCGCGCAATGC 1077
|||
Db 341 LeuGlyProLeuGlnGlySerLeuGlnuArgValGlnuValProAlaGlyGlnuSer 360
OY 1078 ACCAGCGTGCAGAGGCGCTGCGCGCGCGCGCAGCGCTGACCTGACCGCGCTTC 1137
|||
Db 361 ThrThrValGlnGlyLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe 380
OY 1138 CGCTGCGGCGCGGAGAGCGCGCTGTCGCGCAGAGCGCTGACCGCGCGCGCGCGCGCG 1197
|||
Db 381 ArgSerGlyArgGlnArgAlaLeuSerAlaValAlaCysThrAlaSerGlyAlaArgThr 400
OY 1198 CGCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1244
|||
Db 401 Arg--AlaProGlnSerMetArgProGlnuAlaGlyProArgGlnuPro 415
```

RESULT 5

```
US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
```

```

? APPLICANT: Ni et al.
? TITLE OF INVENTION: 52 Human secreted proteins
? FILE REFERENCE: P2043p1
? CURRENT APPLICATION NUMBER: US/09/789,561
? PRIOR APPLICATION NUMBER: PCT/US00/24008
? PRIOR FILING DATE: 2000-08-22
? PRIOR FILING DATE: 2000-08-31
? PRIOR APPLICATION NUMBER: 60/152,317
? PRIOR FILING DATE: 1999-09-03
? PRIOR APPLICATION NUMBER: 60/152,315
? PRIOR FILING DATE: 1999-09-03
? NUMBER OF SEQ ID NOS: 194
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 85
? LENGTH: 215
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (7)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-03-789-561-85

```

Alignment Scores:	
Pred. No.:	1.03e-53
Score:	1067.50
Percent Similarity:	99.1%
Best Local Similarity:	99.1%
Query Match:	44.9%
DB:	3
	Gaps: 1
US-10-639-035A-5 (1-1254) x US-09-789-561-85 (1-215)	

QY	1	ATGCTCCCTCGAAGCGGCGCTGGGCGTGGCCCTGACCTTGCGGCTGGCCCTGCGCCGAGC	60
Db	1	MetLeuProThrPhrHis***GlyLeuAlaLeuSerLeuAlaGlyLeuAlaHisSer	20
QY	61	GCGCGGAGCGCGGCTCCACGAGCATGACGCCCCGAGGGAGCTGATGTTCTCTGTGAC	120
Db	21	GIYALAGIuArgIuYProProAlaSerAlaProArgIuAspLeuMetPheLeuLeuAsp	40
QY	121	AGCTCAGGCCAGGCTCTCTACATACAGAGTTCTCCCGGATTCCGGAGATTGTGGGGACCTG	180
Db	41	SetSerAlaSerValSerHisTryGIuPheSerArgValAArgGIuPheValGIYLeu	60
QY	181	GTCGCTCACAATGGCCCCCTGGGGACCGGGGGCCCTGGCGTGCAGTCTGTGCACGTGGGAGT	240
Db	61	ValAlaProLeuProLeuGIuThrGIuAlaLeuAlaGlyAlaSerLeuValHisValGIYSer	80
QY	241	CGGCGATACACGAGTTCCCTCTTGCGCGACGACGTGGGTGAGGCTGCCAGAGATCG	300
Db	81	ArgProThrGIuThrGIuPheProPheGIuHisSerSerGIYAlaAlaGIuAspAla	100
QY	301	GTCGGTCTTCTGCGCCAGCGGATGGGTGACACCCACATGGCTGGCGCTGGTCTATGCC	360
Db	101	ValArgAlaSerAlaGlnArgMetGIYAspThrHisAlaThrGIYLeuAlaLeuValTYrAla	120
QY	361	AAGGAAGAGTGTTTCTGAGAGATAGAGTGGCGCGGCGACGGGGTSCCAAGATGCTGGT	420
Db	121	LYeGIuGIuLeuPheAlaGluAlaSerGIYAlaArgProGIYAlaProLYuValLeuVal	140
QY	421	TGGGTGACAGATGAGCGGCTTCAGCGCACCTGTGGGCGCCCCCATGACAGAGCTTAAGGAC	480
Db	141	TrpValIhrAspGIYGIYSerSerAspProValGIYProPomeGIuLeuLYuAsp	160
QY	481	CTGGGCGTCAACCGTTTCATTGTCAAGACCGGCGCAGGCACTTCTTGAAGCTGTCAACC	540
Db	161	LeuGIYValIhrValPheIleValSerThrGIYArgGIYAspPheLeuGIuLeuSerAla	180
QY	541	GTGGCGTCAAGCCCTGGCGGAGGAAGCACCTTGAGAGTGGAGATGACCTGCACATC	600
Db	181	AlaAlaSerAlaProAlaGIuLYuHisLYuHisPheValAspValAspAspLeuHisIle	200

```

QY      601 ATTGTCCAGAGCTGAGGGGGCTCATTTCTC---GGATGCGGGCGG 642
        |||||
Db      201 ILevalGInGInLeuArgGlySerIleLeuAspAlaMetArgPro 215

```

```

RESULT 6
US-09-833-245-2096
: Sequence 2096, Application US/09833245
: Publication No. US20040010134A1
: GENERAL INFORMATION:
:   APPLICANT: Human Genome Sciences, Inc.
:   TITLE OF INVENTION: Albumin Fusion Proteins
:   FILE REFERENCE: PF546PCT
:   CURRENT APPLICATION NUMBER: US/09/833,245
:   CURRENT FILING DATE: 2001-04-12
:   PRIOR APPLICATION NUMBER: 60/229, 358
:   PRIOR FILING DATE: 2000-04-12
:   PRIOR APPLICATION NUMBER: 60/256, 911
:   PRIOR FILING DATE: 2000-12-21
:   PRIOR APPLICATION NUMBER: 60/199, 384
:   PRIOR FILING DATE: 2000-04-25
:   NUMBER OF SEQ ID NOS: 2267
:   SOFTWARE: Patentin Ver. 2.1
:   SEQ ID NO 2096
:   LENGTH: 215
:   TYPE: PRT
:   ORGANISM: Homo sapiens
:   FEATURE:
:   NAME/KEY: SITE
:   LOCATION (7)
:   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

```

Alignment Scores:		
Pred. No.:	1, 03e-53	Length: 215
Score:	1067.50	Matches: 213
Percent Similarity:	99.1%	Conservative: 0
Best Local Similarity:	99.1%	Mismatches: 1
Query Match:	44.9%	Indels: 1
DB:	3	Gaps: 1
US-10-699-035A-5 (1-1254) x US-09-833-245-2096 (1-215)		
QY	1 ATGCTCCCTGGACGGCGCTGGCGCTGGACCTTGAGCTTGGCGGCTGGCGGCTGGCGGAGC	60
Db	1 MetLeuProTfThrAla***GlyLeuAlaLeuSerLeuAlaLeuAlaLeuAlaArgSer	20
QY	61 GGGCGGAGCCGGGGTCCACACAGACACACCCCGGAGGGACCTGAGTTCCTGCTGAC	120
Db	21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp	40
QY	121 AGTCAGCAGCGGTCTCTACACAGATTTCCCGGGTCCGGAGTTGTGGGACAGCTG	180
Db	41 SerSerAlaSerValSerHisIleTyrGluPheSerArgValaArgGluPheValGlyGlnLeu	60
QY	181 GTGGCTCACCTGCCCTGGGCAACGGAGGCCCTGGCGTGCAGCTCTGTGACAGTGGAGAGT	240
Db	61 ValAlaProLeuProLeuGlyTyrGlyAlaLeuArgAlaSerLeuValHisValGlySer	80
QY	241 CGGCATACACCGAGTTCCTCTGGGCCACGACAGCTCGGCTAGAGCTGCCACAGAGCG	300
Db	81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla	100
QY	301 GTGCGTGTCTTGGCCAGCGGACATGGGTATGACCCACACATGGCTCGGCGCTGTATGCC	360
Db	101 ValArgAlaSerAlaGlnArgMetCylAspPheHisThrArgLeuAlaLeuValTyrAla	120
QY	361 AAGAAACAGCTGTTCCTGAGACATCAGGTGCCCGGACGGAGGTGCCCAAAGTCTGTG	420
Db	121 LysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyValProLysValLeuVal	140
QY	421 TGGGTATACAGATGGCGGCTCCAGGCAACCTGTGGGCCCCCAATGACAGAGCTCAAGAC	480
Db	141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGluLeuLysASP	160


```

Db      48 GYALAGIARGLYProProAlaSerAlaProArgLYAspLeuMetPheLeuAsp 67
QY      121 ACCTCAGCAGCGGCTCTCTCACTACGAGTTCTCCCGGGTTCCGGAGTTTGCGGCGCAGCTG 180
Db      68 SerSerAlaSerValSerIstYrGIuPheSerArgValArgGIuPheValGIyGIuLeu 87
QY      181 GTGGCTCCACTGCCCCCTGGGCGACCGGGGCGCTGCGTGCAGTGTGTGACGCTGGGCGAGT 240
Db      88 ValAlaProLeuProLeuGIyThrGIyAlaLeuAlaSerIleuValHisValGIySer 107
QY      241 CGGCATACACCGAGTTCCCTTCCGCGACAGCTGCGGGTGGAGGCTGCCAGATGGG 300
Db      108 ArgProIYrThrGIuPheProPheGIyGlnHisSerSerGIyGlnAlaIaGIuAspAla 127
QY      301 GTGGGCTCTTGGCCGAGCGATGGGTGACACCCCACTGGCGCTGGCGCTGTATGGC 360
Db      128 ValArgAlaSerAlaGlnArgMetGIyAspPthHisThrGIyLeuAlaLeuValIYrAla 147
QY      361 AAGGAACAGCTGTTTGCTGAAGCATGAGTGGCCCGGCGACGGGGTGGCCAAAGTGTGTG 420
Db      148 IYSGIuGlnIleuPheAlaGlnIaIaSerGIyAlaArgProGIyValProIYsValIleuVal 167
QY      421 TGGGTGACGATGGCGGCTCTCCAGCAGCCTGTGGGGCCCCCATGTGACGAGCTCAAGAC 480
Db      168 TrpValThrAspGIyGIySerSerAspProValGIyProPheMetGIuIleuIYsAsp 187
QY      481 CTGGGGGTGACCGGTGTTCATTGTGACAGCGGCGGCGGCAACTTCTGGAGCTGTGAGCC 540
Db      188 LeuGIyValThrValPheIleValSerThrGIyAsnPheLeuGIuLeuSerAla 207
QY      541 GGTGCTCAGCCCTGCGCGAGAGCACTGTGACCTTGTGACGCTGTGATGACTGTGACATC 600
Db      208 AlaAlaSerAlaProAlaGlnIYsHisIleuHisPheValAspValAspAspLeuHisIle 227
QY      601 ATTGTCCAAGAGCTGAAGGGCTTCATTCTC---GCCATGGCGCGG 642
Db      228 ILeValGIuGIuLeuArgGIySerIleLeuAspAlaMetArgPro 242

RESULT 9
US-10-883-936-159
; Sequence 159, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: PZ043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; PRIOR FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE

```

```

; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-883-936-159
Alignment Scores:
pred. No.: 1,03e-53 Length: 242
Score: 1067.50 Matches: 213
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 44.9% Indels: 1
DB: 5 Gaps: 1

US-10-699-035a-5 (1-1254) x US-10-883-936-159 (1-242)
QY      1 ATGCTCCCTGAGCGGCGCTCGGCTGCGCTGAGCTTGCGGCTGGCGCGGAGC 60
Db      28 MetLeuProIYrPthAla***GIYleuAlaLeuSerIleuArgLeuAlaIaArgSer 47
QY      61 GGGCGGAGCGCGGTCCACAGCATCAGCCCCCGAGGGGAGCTGATGTTCTGCTGGAC 120
Db      48 GYALAGIARGLYProProAlaSerAlaProArgGIYAspLeuMetPheLeuAsp 67
QY      121 ACCTCAGCAGCGTCTCTCACTACGAGTTCTCCCGGGTTCCGGAGTTTGCGGCGCAGCTG 180
Db      68 SerSerAlaSerValSerIstYrGIuPheSerArgValArgGIuPheValGIyGIuLeu 87
QY      181 GTGGCTCCACTGCCCCCTGGGCGACCGGGGCGCTGCGTGCAGTGTGTGACGCTGGGCGAGT 240
Db      88 ValAlaProLeuProLeuGIyThrGIyAlaLeuAlaSerIleuValHisValGIySer 107
QY      241 CGGCATACACCGAGTTCCCTTCCGCGACAGCTGCGGGTGGAGGCTGCCAGATGGG 300
Db      108 ArgProIYrThrGIuPheProPheGIyGlnHisSerSerGIyGlnAlaIaGIuAspAla 127
QY      301 GTGGGCTCTTGGCCGAGCGATGGGTGACACCCCACTGGCGCTGGCGCTGTATGGC 360
Db      128 ValArgAlaSerAlaGlnArgMetGIyAspPthHisThrGIyLeuAlaLeuValIYrAla 147
QY      361 AAGGAACAGCTGTTTGCTGAAGCATGAGTGGCCCGGCGACGGGGTGGCCAAAGTGTGTG 420
Db      148 IYSGIuGlnIleuPheAlaGlnIaIaSerGIyAlaArgProGIyValProIYsValIleuVal 167
QY      421 TGGGTGACGATGGCGGCTCTCCAGCAGCCTGTGGGGCCCCCATGTGACGAGCTCAAGAC 480
Db      168 TrpValThrAspGIyGIySerSerAspProValGIyProPheMetGIuIleuIYsAsp 187
QY      481 CTGGGGGTGACCGGTGTTCATTGTGACAGCGGCGGCGGCAACTTCTGGAGCTGTGAGCC 540
Db      188 LeuGIyValThrValPheIleValSerThrGIyAsnPheLeuGIuLeuSerAla 207
QY      541 GGTGCTCAGCCCTGCGCGAGAGCACTGTGACCTTGTGACGCTGTGATGACTGTGACATC 600
Db      208 AlaAlaSerAlaProAlaGlnIYsHisIleuHisPheValAspValAspAspLeuHisIle 227
QY      601 ATTGTCCAAGAGCTGAAGGGCTTCATTCTC---GCCATGGCGCGG 642
Db      228 ILeValGIuGIuLeuArgGIySerIleLeuAspAlaMetArgPro 242

RESULT 10
US-09-789-561-165
; Sequence 165, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: PZ043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315

```

```
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 165
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-789-561-165

Alignment Scores:
Pred. No.: 2,14e-53 Length: 226
Score: 1062.00 Matches: 210
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 44.6% Indels: 0
DB: 3 Gaps: 0

US-10-699-035A-5 (1-1254) x US-09-789-561-165 (1-226)

QY 1 ATGCTCCCTGGACGGCGCTCGGCTGAGCTTGGCGCTGGCGCGGAGC 60
Db 14 MetLeuProTPrThAlaLeuGlYleuAlaLeuSerLeuAlaLeuAlaArgSer 33
QY 61 GCGCGGAGCGCGGTCCACAGATCAGCCCCGAGGGAGCTGATGTTCTGCTGAC 120
Db 34 GlYAlaGlAArgGlYProProAlaSerAlaProAlaArgYAlaAspLeuMetPheLeuLeuAsp 53
QY 121 AGCTCAGCAGCGCTCTCTACACGAGTTCTCCGGGTTCCGGAGTTTGTGGGAGCTG 180
Db 54 SerSerAlaSerValSerHisTyrGluPheSerAlaArgGluPheValGlyGlnLeu 73
QY 181 GTGCTCCACTGCCCCCTGGGCAACCGGGCCCTGCGTGCAGTCTGCTGACGCTGGGCACT 240
Db 74 ValAlaProLeuProLeuGlYThrGlyAlaLeuAlaArgAlaSerLeuValHisValGlySer 93
QY 241 CGGCGATACACCGAGTTCCCTTGGCCGACAGCTGGGTGAGGCTGCCAGAGATGG 300
Db 94 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla 113
QY 301 GTGCGTGTCTTGGCCAGCGACATGGGTGACACCCACACTGGCGCTGGCGCTTATGCC 360
Db 114 ValAlaGlaSerAlaGlnIleArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 133
QY 361 AAGAACAGCTGTTTGTCTGAAGCATCAGGTGCCCGGCGAGGGGTGCCAAAGTGTGTTG 420
Db 134 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProYlsvAlleuVal 153
QY 421 TGGGTGACAGATGGGGCTCCAGCGACCTGTGGGCCCCCATGAGGAGGCTCAAGGAC 480
Db 154 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnIleuLeuYAsp 173
QY 481 CTGGCGCTCAGCGCTTCTCATTTGTACAGACCGGCGAGGCAACTTCTTGAGCTGTACGC 540
Db 174 LeuGlyValThrValPheIleValSerThrGlyAsnPheLeuGluLeuSerAla 193
QY 541 GCTGCTCAGCCCTTCCGAGAAAGCACTTGACCTTGTGACGTGATGACCTGCACATC 600
Db 194 AlaAlaSerAlaProAlaGlnIleYshIleuHisPheValAspValAspAspLeuHisIle 213
QY 601 ATTGTCGAAGAGCTGAGGGCTTCATTCTC 630
Db 214 IleValGlnGluLeuArgGlySerIleLeu 223

RESULT 11
US-10-883-936-165
; Sequence 165, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043PI
; CURRENT APPLICATION NUMBER: US/10/883,936
; CURRENT FILING DATE: 2004-07-06
```

```
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 165
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-883-936-165

Alignment Scores:
Pred. No.: 2,14e-53 Length: 226
Score: 1062.00 Matches: 210
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 44.6% Indels: 0
DB: 5 Gaps: 0

US-10-699-035A-5 (1-1254) x US-10-883-936-165 (1-226)

QY 1 ATGCTCCCTGGACGGCGCTCGGCTGAGCTTGGCGCTGGCGCGGAGC 60
Db 14 MetLeuProTPrThAlaLeuGlYleuAlaLeuSerLeuAlaLeuAlaArgSer 33
QY 61 GCGCGGAGCGCGGTCCACAGATCAGCCCCGAGGGAGCTGATGTTCTGCTGAC 120
Db 34 GlYAlaGlAArgGlYProProAlaSerAlaProAlaArgYAlaAspLeuMetPheLeuLeuAsp 53
QY 121 AGCTCAGCAGCGTCTCTACACGAGTTCTCCGGGTTCCGGAGTTTGTGGGAGCTG 180
Db 54 SerSerAlaSerValSerHisTyrGluPheSerAlaArgGluPheValGlyGlnLeu 73
QY 181 GTGCTCCACTGCCCCCTGGGCAACCGGGCCCTGCGTGCAGTCTGCTGACGCTGGGCACT 240
Db 74 ValAlaProLeuProLeuGlYThrGlyAlaLeuAlaArgAlaSerLeuValHisValGlySer 93
QY 241 CGGCGATACACCGAGTTCCCTTGGCCGACAGCTGGGTGAGGCTGCCAGAGATGG 300
Db 94 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla 113
QY 301 GTGCGTGTCTTGGCCAGCGACATGGGTGACACCCACACTGGCGCTGGCGCTTATGCC 360
Db 114 ValAlaGlaSerAlaGlnIleArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 133
QY 361 AAGAACAGCTGTTTGTCTGAAGCATCAGGTGCCCGGCGAGGGGTGCCAAAGTGTGTTG 420
Db 134 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProYlsvAlleuVal 153
QY 421 TGGGTGACAGATGGGGCTCCAGCGACCTGTGGGCCCCCATGAGGAGGCTCAAGGAC 480
Db 154 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnIleuLeuYAsp 173
QY 481 CTGGCGCTCAGCGCTTCTCATTTGTACAGACCGGCGAGGCAACTTCTTGAGCTGTACGC 540
Db 174 LeuGlyValThrValPheIleValSerThrGlyAsnPheLeuGluLeuSerAla 193
QY 541 GCTGCTCAGCCCTTCCGAGAAAGCACTTGACCTTGTGACGTGATGACCTGCACATC 600
Db 194 AlaAlaSerAlaProAlaGlnIleYshIleuHisPheValAspValAspAspLeuHisIle 213
QY 601 ATTGTCGAAGAGCTGAGGGCTTCATTCTC 630
Db 214 IleValGlnGluLeuArgGlySerIleLeu 223

RESULT 12
US-09-789-561-160
; Sequence 160, Application US/09789561
```

Patent No. US20020064818A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: P2043P1
CURRENT APPLICATION NUMBER: US/09/789,561
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 160
LENGTH: 186
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (152)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (180)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (184)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (186)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-160

Alignment Scores:
Pred. No.: 1,73e-44 Length: 186
Score: 908.00 Matches: 156
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 38.2% Indels: 0
DB: Gaps: 0

US-10-699-035A-5 (1-1254) x US-09-789-561-160 (1-186)
QY 472 GCTCTGATGGGGGGGGCCACAGGAGGCTGCTGAGCGCATGTGTACCCACACGACA 413
DB 1 AAlaProAlaTrpGlyGlyProGlnGlyArgTrpSerArgHisLeuSerProThrProAla 20
QY 412 CTTTGGGACCCCTGGCCGGGACCTGATGCTTACAGCAACAGCTGTTCTTGGCATAGA 353
DB 21 LeuTrpAlaProLeuAlaGlyHisLeuMetLeuGlnGlnThrAlaValProThrPheArg 40
QY 352 CCAAGCCGACGAGGCTGAGTGGGTGTACCCATGCGCTGGGCGAGAAAGCAGCAGCATCCT 293
DB 41 ProAlaProGlyGlnGlyGlyGlyCysHisProCysAlaGlyGlnGlyHisAlaProHisPhe 60
QY 292 GGGCAGCTCAGCCGAGGCTGCTGCTGGCCGAGGGAAGTGGTGTATGGCCGACCTGCCA 233
DB 61 GlyGlnProHisProSerCysAlaGlyArgGlyThrArgCysMetAlaSerCysPhe 80
QY 232 CGTGCACAGACTGGACGACGAGGCGCCGGTGGCCAGGGGACAGTGAAGCCACGAGTGC 173
DB 81 ArgAlaProAspTrpHisAlaGlyProArgCysPheGlyAlaValGlnProProAlaAla 100
QY 172 CCAGAACTCCGACACCCGGGAGAACTGTATGAGAGAGAGCTGGCTGAGCTGTCCAGCA 113
DB 101 ProGlnThrProGlnProGlyArgTrpArgSerGlnArgTrpLeuSerCysProAla 120
QY 112 GGAACATCAGGCTCCCTCGGGGGGCTGATGCTGTGAGACGCGCTCGGCGCGCTCGCG 53
DB 121 GlyThrSerGlyProLeuGlyGlyLeuMetLeuValAlaSerArgAlaProAlaArgSerAla 140
QY 52 CCAAGCCGACGAGCTCAGGCGCAGGCGGAGCGCGCTCCAGGGGAGCA 2

DB 141 ProAlaProAlaAlaSerGlyProGlyArg***ProSerArgGlyAla 157

RESULT 13

US-10-883-936-160

Sequence 160, Application US/10883936
Publication No. US2005001986A1
GENERAL INFORMATION:

APPLICANT: Ni et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: P2043P1
CURRENT APPLICATION NUMBER: US/10/883,936
CURRENT FILING DATE: 2004-07-06
PRIOR APPLICATION NUMBER: US/09/789,561
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 160
LENGTH: 186
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (152)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (180)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (184)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (186)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-160

Alignment Scores:
Pred. No.: 1,73e-44 Length: 186
Score: 908.00 Matches: 156
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 38.2% Indels: 0
DB: Gaps: 0

US-10-699-035A-5 (1-1254) x US-10-883-936-160 (1-186)
QY 472 GCTCTGATGGGGGGGGCCACAGGAGGCTGCTGAGCGCATGTGTACCCACACGACA 413
DB 1 AAlaProAlaTrpGlyGlyProGlnGlyArgTrpSerArgHisLeuSerProThrProAla 20
QY 412 CTTTGGGACCCCTGGCCGGGACCTGATGCTTACAGCAACAGCTGTTCTTGGCATAGA 353
DB 21 LeuTrpAlaProLeuAlaGlyHisLeuMetLeuGlnGlnThrAlaValProThrPheArg 40
QY 352 CCAAGCCGACGAGGCTGAGTGGGTGTACCCATGCGCTGGGCGAGAAAGCAGCAGCATCCT 293
DB 41 ProAlaProGlyGlnGlyGlyGlyCysHisProCysAlaGlyGlnGlyHisAlaProHisPhe 60
QY 292 GGGCAGCTCAGCCGAGGCTGCTGCTGGCCGAGGGAAGTGGTGTATGGCCGACCTGCCA 233
DB 61 GlyGlnProHisProSerCysAlaGlyArgGlyThrArgCysMetAlaSerCysPhe 80
QY 232 CGTGCACAGACTGGACGACGAGGCGCCGGTGGCCAGGGGACAGTGAAGCCACGAGTGC 173

Db 81 ArgAlaProAbgETrpHisAlaGlyProAbgCySPProGlyAlaValGluProAlaAla 100
Qy 172 CCACAACCTCCCGAACCAGGAGAACTGTAAGTGAAGAGCGCTGAGCTGTCACGA 113
Db 101 ProGlnThrProGluProGlyArgThrArgSerGluArgArgTrpLeuSerCySPProAla 120
Qy 112 GGAACATCAGGTCCTCCCTCGGGGGGCTGATGCTGTGTGACCGCGCTCCGCGCTCCGG 53
Db 121 GlyThrSerGlyProLeuGlyGlyLeuMetLeuValAspArgAlaProArgHisSerAla 140
Qy 52 CCAGGCGCAGCCGCAAGCTCAGGCGCAGGCGCGCTCCAGGGGAGCA 2
Db 141 ProAlaProAlaAlaSerSerCyProGlyArg***ProSerArgGlyAla 157
RESULT 14
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2
Alignment Scores:
Pred. No.: 1,97e-44 Length: 180
Score: 907.00 Matches: 179
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 38.1% Indels: 0
DB: 4 Gaps: 0
US-10-699-035A-5 (1-1254) x US-10-699-035A-2 (1-180)
Qy 94 CGAGGGAGCTATGTTCTCTGCTGACACTCAGCCAGGCTCTTCTCAGAGTTTCTCC 153
Db 1 ArgGlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyGluPheSer 20
Qy 154 CGGTTTCGGAGATTGTTGGGAGAGCTGTGCTCAGCTCCCTGGGAGCCGGGGCCCTG 213
Db 21 ArgValAlaArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeu 40
Qy 214 CGTGCAGTCTGTGTGACGTGGGCACTGGCCCATACACCGAGTCTCCCTTCGGCCAGCAC 273
Db 41 ArgAlaSerLeuValHisValGlySerArgProTyThrGluPheProPheGlyGlnHis 60
Qy 274 AGCTGGGTGAGAGCTGCCAGAGATGGGTCGCTTTCGCCAGGGCATTTGGGTACACC 333
Db 61 SerSerGlyGlnAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThr 80
Qy 334 CACACTGGCGCTGGCGCTGCTATGSCCAAGAAAGCTTTTGCTGAACATCAGGTGCC 393
Db 81 HisThrGlyLeuAlaLeuValItyrAlaGlyGlnLeuPheAlaGlnAlaSerGlyAla 100
Qy 394 CGGCGAGGGGTGCCCAAGTGTGTGTGTGTGTGACAGATGGCGGCTCCAGCACTGTG 453
Db 101 ArgProGlyValProGlyValLeuValItyrValIthAspGlyGlySerSerAspProVal 120
Qy 454 GGGCCCCCATGACAGAGCTCAAGAGCTGGGGCGTACCGGTTCATTGTACGACCGGC 513

Db 121 GlyProPheMetGlnGluLeuLeuAspLeuGlyValIthrValPheIleValSerThrGly 140
Qy 514 CGAGGCAACTTCTGAGACTGTCAAGCCGCTGCTCAGCCCTGCGGAGAACACTGCAC 573
Db 141 ArgGlyAsnPheLeuGlnLeuSerAlaAlaIlaSerAlaProAlaGlnGlyHisIleuHis 160
Qy 574 TTTGTGACGTGATGACCTGCACATCATTTGTCCAGAAGCTGAGGGGCTCCATTCTC 630
Db 161 PheValAspValaAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleu 179
RESULT 15
US-10-149-819-18
; Sequence 18, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyoung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 18
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CD1
US-10-149-819-18
Alignment Scores:
Pred. No.: 4.1e-44 Length: 185
Score: 901.50 Matches: 180
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 0
Query Match: 37.9% Indels: 1
DB: 4 Gaps: 1
US-10-699-035A-5 (1-1254) x US-10-149-819-18 (1-185)
Qy 106 ATGTTCTCGCTGAGACAGTCAAGCCAGGCTCTCTCAGATTCTCCGGGTTTCGGAG 165
Db 1 MetPheLeuLeuAspSerSerAlaSerValSerHisTyGlyGluPheSerArgValArgGlu 20
Qy 166 TTTGTGGGAGCTGTGTGCTCACTGCCCTTGAGGCAACGGGGCCCTCGTGCACAGTCTG 225
Db 21 PheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerIleu 40
Qy 226 GTGCAGTGGGAGTGGCCATPACCGAGTTTCCCTTGGGCGCACACAGCTCGGGTACG 285
Db 41 ValHisValGlySerArgProTyThrGluPheProPheGlyGlnHisSerSerGlyGlu 60
Qy 286 GCTGCCAGAGATGCGGTCGCTTTCGCCAGCGGAGGAGTGCACCCAGATGGCGCTG 345
Db 61 AlaIlaGlnAspAlaValaArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeu 80
Qy 346 GCGCTGCTCTATGCAAGAAAGAGCTTTTGTGCTGAAGCATAGGTGCCGGGAGGGGTG 405
Db 81 AlaLeuValItyrAlaGlyGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyVal 100

Qy	406	CCCAAGTCGTGCTGGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCCCCCAGT	465
Db	101	ProlyValIleuValItrValIthrAspGlyGlySerSerAspProValGlyProProme	120
Qy	466	CAGAGCTCAAGAGCCTGGGCGTCAACGTTGATTCAGACACCGCGAGGCACTTC	525
Db	121	GInGluIleuIleuAspIleuGlyValIthrValIleuValSerThrGlyArgGlyAsn	140
Qy	526	CTGAGCTGTCAAGCGCTGCTCAGCCCTGCGAGAGACACCTGCACCTTTGTGACGTG	585
Db	141	LeuGluIleuSerAlaIleuValSerAlaProIleuGluIleuHisIleuValAspVal	160
Qy	586	GATGACCTGCACATCTTGTCCAGAGCTGAGGGGCTCCATTCTC--GCGATGGCGCG	642
Db	161	AspAspIleuHisIleuValGInGluIleuArgGlySerIleuAspAlaMetArgPro	180
Qy	643	CAG 645	
Db	181	Gln 181	

Search completed: February 13, 2006, 13:53:31
 Job time : 212.799 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:33:34 ; Search time 109.565 Seconds

(without alignments)
721.837 Million cell updates/sec

Title: US-10-699-035A-2

Sequence: 1 RGLMFLLSASASVSHYFS.....FVDDVLLHIVQLRSLTD 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 segs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_21.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003s.*
7: geneseqp2003bs.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	913	100.0	180	6	AAE32500 Human von
2	913	100.0	215	4	AAE87344 Human gen
3	913	100.0	215	5	ABG65347 Human alb
4	913	100.0	215	8	ADL78614 Human alb
5	913	100.0	226	4	AAE87424 Human gen
6	913	100.0	242	4	AAE87418 Human gen
7	913	100.0	285	4	AAE88340 Human mem
8	913	100.0	285	9	ADY63045 Human cto
9	913	100.0	418	6	AAE32502 Human wil
10	913	100.0	445	5	ABP69674 Human pol
11	913	100.0	445	8	ADH71106 Human pro
12	892	97.7	185	4	AAE03654 Human ext
13	883	96.7	299	3	AAE82581 Human ORF
14	768	84.1	180	6	AAE32503 Human ORF
15	768	84.1	415	6	AAE32501 Mouse von
16	761	83.4	421	2	AAE66326 Kidney in
17	384	42.1	77	3	AAE02196 Human sec
18	272.5	29.8	176	5	ABE78816 Von Wille
19	265	29.0	644	8	ADR66963 Mouse can
20	265	29.0	644	8	ADZ12632 Murine ca
21	263.5	28.9	3063	5	ABE90762 Human Tum
22	263.5	28.9	3063	6	ABU54469 Human tum
23	263.5	28.9	3063	6	ABR47415 Breast ca
24	263.5	28.9	3063	6	ABR47416 Breast ca

25	263.5	28.9	3063	8	ADJ75666 Marker ge
26	263.5	28.9	3118	4	AAU27790 Human full
27	259	28.4	293	3	AAE39142 Human sec
28	259	28.4	357	4	AAE93189 Human pol
29	259	28.4	357	8	ADJ30530 Human pro
30	259	28.4	391	9	ADZ12639 Human can
31	259	28.4	488	8	AAE93707 Human can
32	259	28.4	488	8	ADL31608 Human pol
33	259	28.4	499	9	ADZ12645 Human can
34	259	28.4	499	9	ADZ12643 Human can
35	259	28.4	540	9	ADZ12637 Human can
36	259	28.4	540	9	ADZ12641 Human can
37	259	28.4	580	8	ADQ19949 Human can
38	259	28.4	581	8	ADR66966 Human can
39	259	28.4	581	9	ADZ12645 Human can
40	259	28.4	581	9	ADZ12647 Human can
41	259	28.4	620	9	ADZ12649 Human can
42	256.5	28.1	755	5	AAU84267 Human end
43	256.5	28.1	795	4	AAE27229 Human ECM
44	256.5	28.1	1780	7	ADE15980 G-coupled
45	256.5	28.1	1780	8	ADL93919 Human G-C

ALIGNMENTS

RESULT 1	AAE32500 standard; protein; 180 AA.
ID	AAE32500
AC	AAE32500;
DT	24-MAR-2003 (first entry)
DE	Human von Willebrand Factor A (VA) domain.
KW	Willebrand Factor A domain related-protein; von Willebrand Factor A; VA; extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP; gene therapy; human.
OS	Homo sapiens.
FT	Misc-difference 1..2
FT	/note="Encoded by GCG"
FT	/note="179..180
FT	/note="Encoded by CTC"
PN	WO200288184-A1.
PD	07-NOV-2002.
PF	02-MAY-2002; 2002MO-AU000542.
PR	02-MAY-2001; 2001AU-00004701.
PA	(MURD-) MURDOCH CHILDRENS RES INST.
PI	Bateman JF, Fitzgerald DJ;
DR	WPI, 2003-111873/10.
DR	N-PSDB; AAD50397.
PT	New isolated Willebrand Factor A-Related Protein polypeptide useful for the manufacture of a medicament in the treatment of a disease condition of the extracellular matrix, in particular arthritis.
PT	Claim 7, Page 72-73; 103pp; English.
PS	The invention relates to Willebrand Factor A domain related-protein (WARP) which is a member of von Willebrand Factor A (VA)-domain protein superfamily of extracellular matrix (ECM) proteins. WARP is used as a molecular marker, used for detecting a loss of ECM integrity in an animal

CC subject, monitoring repair, regeneration or other disease processes in an
 CC animal subject and detecting a disease condition or a propensity for the
 CC development of a disease condition in an animal subject. The invention is
 CC useful for the manufacture of a medicament in the treatment of a disease
 CC condition of the ECM. The disease condition involves the cartilage, and
 CC is preferably arthritis. The invention is also used in gene therapy. The
 CC present sequence is human VA domain

XX Sequence 180 AA;

Query Match 100.0%; Score 913; DB 6; Length 180;
 Best Local Similarity 100.0%; Pred. No. 1.6e-94;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLTGALPASLVHVGSRPYTEFPFGQH 60
 DB 1 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLTGALPASLVHVGSRPYTEFPFGQH 60
 QY 61 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQLFAEASGARGPVKVLWMTDGGSSDPV 120
 DB 61 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQLFAEASGARGPVKVLWMTDGGSSDPV 120
 QY 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 180
 DB 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 180

RESULT 2

AAAB87344
 ID AAB87344 standard; protein; 215 AA.

XX AAB87344;

DT 22-MAY-2001 (first entry)

DB Human gene 3 encoded secreted protein HMTB078, SEQ ID NO:85.

XX Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnery; cell culture;
 KW chemotaxis; food additive; binding partner identification.

XX Homo sapiens.

XX WO200118022-A1.

XX 15-MAR-2001.

XX 31-AUG-2000; 2000WO-US024008.

XX 03-SEP-1999; 99US-0152315P.

XX 03-SEP-1999; 99US-0152317P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Baker KP, Birse CE, Fiacella M, Komatsoulis GA, Rosen CA;
 PI Sopet DR, Young PE, Eshner R, Duan DR, Olsen HS, Lafleur DW,
 PI Moore PA, Shi Y, Wei Y, Florence KA;

XX WPI; 2001-203081/20.

XX N-PSDB; AAF91860.

PT Nucleic acid molecules encoding human secreted proteins, used in
 PT treating, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers.

PS Claim 11; Page 532-533; 607p; English.

XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
 CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
 CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 52 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
 CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention

XX Sequence 215 AA;

Query Match 100.0%; Score 913; DB 4; Length 215;
 Best Local Similarity 100.0%; Pred. No. 2e-94;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLTGALPASLVHVGSRPYTEFPFGQH 60
 DB 32 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLTGALPASLVHVGSRPYTEFPFGQH 91
 QY 61 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQLFAEASGARGPVKVLWMTDGGSSDPV 120
 DB 92 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQLFAEASGARGPVKVLWMTDGGSSDPV 151
 QY 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 180
 DB 152 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 211

RESULT 3

ABG65347
 ID ABG65347 standard; protein; 215 AA.

XX ABG65347;

DT 27-AUG-2002 (first entry)

DE Human albumin fusion protein #2022.

XX Albumin fusion protein, therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective tissue disorder;
 KW cytoskeletal; antifertility; antiinflammatory; anticancer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.

XX Homo sapiens.
 OS Synthetic.

PN WO200177137-A1.

XX 18-OCT-2001.
 PD 12-APR-2001; 2001WO-US011988.
 XX PF 12-APR-2001; 2000US-0229358P.
 XX PR 12-APR-2000; 2000US-0229358P.
 XX PR 25-APR-2000; 2000US-0199384P.
 XX PR 21-DEC-2000; 2000US-0256931P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Haseltine WA;
 PI WPI; 2002-010886/01.
 DR New fusion protein for treating disease e.g. diabetes comprises an
 XX albumin fused to a therapeutic protein.
 PT Claim 1; Page 1935; 2102pp; English.
 PS The present invention relates to albumin fusion proteins comprising a
 XX albumin, HSA). The proteins are useful for treating a disease or disorder
 CC that may be modulated by therapeutic protein X. The albumin extends the
 CC shelf-life of protein X, and may increase its biological in vitro/in vivo
 CC activity. The protein is useful for treating and diagnosing disorders
 CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
 CC disease, ulcerative colitis), immune disorders (e.g. acquired
 CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
 CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,
 CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
 CC schizophrenia), and connective tissue disorders (e.g. osteoporosis, arthritis).
 CC ABG63326-ABG65518 represent albumin fusion proteins of the invention
 XX
 SQ Sequence 215 AA;
 Query Match 100.0%; Score 913; DB 5; Length 215;
 Best Local Similarity 100.0%; Pred. No. 2e-94;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RDDMFLDSSASVSHYERSRREPFGQVAPLPCTGTLRLSLVHVGSRPTTEPPFGQH 60
 DB 32 RDDMFLDSSASVSHYERSRREPFGQVAPLPCTGTLRLSLVHVGSRPTTEPPFGQH 91
 QY 61 SSGEAAQDVRASAOBMGDTHGLAVYAKEDLPFAASGAPGVPKVLVWMTDGGSSDPV 120
 DB 92 SSGEAAQDVRASAOBMGDTHGLAVYAKEDLPFAASGAPGVPKVLVWMTDGGSSDPV 151
 QY 121 GPMQELKDLGYTVFIVSTGRGNFLELSAASAPAEKHLFPVVDVLDLHIIIVQLRSLID 180
 DB 152 GPMQELKDLGYTVFIVSTGRGNFLELSAASAPAEKHLFPVVDVLDLHIIIVQLRSLID 211
 RESULT 4
 ADL78614 ID ADL78614 strand; protein; 215 AA.
 AC ADL78614;
 XX 20-MAY-2004 (first entry)
 DE Albumin fusion protein related therapeutic protein X, SEQ ID NO 2096.
 XX
 KW albumin fusion protein; cytostatic; antianaemic; antiarthritic;
 KW antiaesthetic; anti-HIV; immunosuppressive; antiinflammatory;
 KW antipsoptic; antibacterial; osteopathic; dermatological; antitumor;
 KW immunomodulator; antiarrhythmic; cardiac; neurotropic; antiipaeamic;
 KW nephrotoxic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
 KW antidiabetic; anabolic; hypertensive; vulnery; gene therapy; cancer;
 KW reproductive system disorder; therapeutic protein.
 XX
 OS Unidentified.
 XX

PN US2004010134-A1.
 XX 15-JAN-2004.
 XX 12-APR-2001; 2001US-00833245.
 XX PF 12-APR-2001; 2000US-0229358P.
 XX PR 12-APR-2000; 2000US-0229358P.
 XX PR 25-APR-2000; 2000US-0199384P.
 XX PR 21-DEC-2000; 2000US-0256931P.
 XX (ROSE/) ROSEN C A.
 PA (HASE/) HASELTINE W A.
 PI Rosen CA, Haseltine WA;
 DR WPI; 2004-090519/09.
 XX New albumin fusion proteins, useful for diagnosing, treating, preventing
 PT or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,
 PT asthma, inflammatory bowel disease or Alzheimer's disease.
 PS Disclosure; SEQ ID NO 2096; 279pp; English.
 XX The invention relates to a novel albumin fusion protein. The invention
 CC further relates to: a composition comprising the albumin fusion protein
 CC and a pharmaceutical carrier; a kit comprising the composition of the
 CC albumin fusion protein formula; a method of treating a disease or
 CC disorder in a patient comprising the step of administering the albumin
 CC fusion protein; a method of treating a patient with a disease or disorder
 CC that is modulated by therapeutic protein X, or its fragment or variant;
 CC a method of extending the shelf life of therapeutic protein X, or its
 CC fragment or variant; a nucleic acid molecule comprising a polynucleotide
 CC sequence encoding the albumin fusion protein; a vector comprising the
 CC nucleic acid molecule of the albumin fusion protein; and a host cell
 CC comprising the nucleic acid molecule of the albumin fusion protein. The
 CC albumin fusion protein and its compositions have the following
 CC activities: cytostatic, antianaemic, antiarthritic, antiaesthetic, anti-
 CC HIV, immunosuppressive, antiinflammatory, antipsoptic, antibacterial,
 CC osteopathic, dermatological, antitumor, immunomodulator, antiarrhythmic,
 CC cardiac, neurotropic, antiipaeamic, nephrotoxic, uropathic,
 CC neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic,
 CC hypertensive, and vulnery. The albumin fusion protein nucleic acid may
 CC be used in gene therapy to treat disorders. The albumin fusion protein is
 CC useful for diagnosing, treating, preventing or ameliorating diseases or
 CC disorders comprising indication: Y. The diseases or disorders include:
 CC cancer (e.g. leukemia, colon, bone, breast, liver or lung cancer),
 CC immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute
 CC lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS),
 CC autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
 CC disease), reproductive system disorders (e.g. prostatitis, inguinal
 CC hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
 CC leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,
 CC Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
 CC or cachexia), cardiovascular disease (e.g. Thrombomias, heart disease,
 CC arrhythmia, cardiac arrest, heat valve disease, hypernatremia or
 CC hyponatraemia), mixed foetal diseases (e.g. foetal alcohol syndrome,
 CC Down's syndrome, Parau syndrome, Turner's syndrome, Apert syndrome or Tay-
 CC Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
 CC tract infections or renal disorders), neural or sensory disease (e.g.
 CC Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
 CC cerebellar ataxia, attention deficit disorder, autism or obsessive
 CC compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
 CC occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
 CC disease or glomerulonephritis), digestive diseases (e.g. portal
 CC hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
 CC or connective tissue or epithelial diseases (e.g. Crohn's disease,
 CC scleroderma, wound healing or epidermolysis bullosa). This sequence
 CC represents a therapeutic protein X relating to the albumin fusion protein
 CC of the invention. The sequence listing data for this specification was
 CC downloaded from the USPTO website.
 XX
 SQ Sequence 215 AA;
 XX

Query Match 100.0%; Score 913; DB 8; Length 215;
Best Local Similarity 100.0%; Pred. No. 2e-94;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RGDLMFLDSSASVSHYEFSSRREVFQGLVAPLPLGTGALRASLVHVSRRPYTFEPFGQH 60
Db 32 RGDLMFLDSSASVSHYEFSSRREVFQGLVAPLPLGTGALRASLVHVSRRPYTFEPFGQH 91
Oy 61 SSGEAAQDAVAVASQRMGDTHTGTLVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 120
Db 92 SSGEAAQDAVAVASQRMGDTHTGTLVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 151
Oy 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
Db 152 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211

RESULT 5
AAB87424
ID AAB87424 standard; protein; 226 AA.
XX
AC AAB87424;
XX
XX 22-MAY-2001 (first entry)
XX
XX Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerability; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200118022-A1.
XX
PD 15-MAR-2001.
XX
PF 31-AUG-2000; 2000WO-US024008.
XX
PR 03-SEP-1999; 99US-0152315P.
PR 03-SEP-1999; 99US-0152317P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Fiascelia M, Komatsoulis GA, Rosen CA;
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
DR WPI; 2001-203081/20.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX
PS Disclosure; Page 18; 607P; English.
XX
CC AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AA87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of

CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
SQ Sequence 226 AA;

Query Match 100.0%; Score 913; DB 8; Length 226;
Best Local Similarity 100.0%; Pred. No. 2.2e-94;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RGDLMFLDSSASVSHYEFSSRREVFQGLVAPLPLGTGALRASLVHVSRRPYTFEPFGQH 60
Db 45 RGDLMFLDSSASVSHYEFSSRREVFQGLVAPLPLGTGALRASLVHVSRRPYTFEPFGQH 104
Oy 61 SSGEAAQDAVAVASQRMGDTHTGTLVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 120
Db 105 SSGEAAQDAVAVASQRMGDTHTGTLVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 164
Oy 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
Db 165 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 224

RESULT 6
AAB87418
ID AAB87418 standard; protein; 242 AA.
XX
AC AAB87418;
XX
XX 22-MAY-2001 (first entry)
XX
XX Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerability; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200118022-A1.
XX
PD 15-MAR-2001.
XX
PF 31-AUG-2000; 2000WO-US024008.
XX
PR 03-SEP-1999; 99US-0152315P.
PR 03-SEP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX

PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA,
PI Soppet DR, Young PG, Ebner R, Duan DR, Olsen HS, Lafleur DW,
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX MPI: 2001-203081/20.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX Disclosure: Page 18; 607pp; English.
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87442-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX SQ Sequence 242 AA;
XX
XX Query Match 100.0%; Score 913; DB 4; Length 242;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFRRVREFVGVLPPLGTGALRASLVHVGSRPYTEPFPGQH 60
XX |||||
XX 59 RGDLMFLDSSASVSHYEFRRVREFVGVLPPLGTGALRASLVHVGSRPYTEPFPGQH 118
XX |||||
XX QY 61 SSGEAAQDAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 120
XX |||||
XX DB 119 SSGEAAQDAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 178
XX |||||
XX QY 121 GPPMDELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDLHIIVOELKGSITLD 180
XX |||||
XX DB 179 GPPMDELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDLHIIVOELKGSITLD 238
XX |||||
XX
XX RESULT 7
XX ID AAB88340 standard; protein; 285 AA.
XX
XX AAB88340;
XX
XX 23-MAY-2001 (first entry)
XX
XX Human membrane or secretory protein clone PSRC0053.
XX
XX Human; secretory protein; membrane protein; vaccine; gene therapy;
XX
XX Rheumatoid arthritis; diabetes.
KW

XX XX Homo sapiens.
XX OS
XX EPI067182-A2.
XX PN
XX 10-JAN-2001.
XX
XX PD
XX 07-JUL-2000; 2000EP-00114090.
XX
XX PF
XX 08-JUL-1999; 99JP-00194179.
XX PR 11-JAN-2000; 2000JP-00118775.
XX PR 02-MAY-2000; 2000JP-00183766.
XX
XX PA (HELI-1) HELIX RES INST.
XX
XX PI Ota T, Isega T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
XX DR MPI: 2001-093989/11.
XX
XX DR N-PSDB; AAF93767.
XX
XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX gene therapy or as candidate target molecules in drug development.
XX
XX PS Claim 1; SEQ ID NO 48; 609pp + Sequence Listing; English.
XX
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX which encode human secretory or membrane proteins represented by AAB88317
XX - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
XX AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
XX invention. The invention also includes methods for the production of
XX antibodies directed against the proteins, and cDNA sequences, which can
XX be used in vaccines. The polynucleotide sequences can be used in gene
XX therapy. The polynucleotide sequences and the proteins they encode may be
XX used in the prevention, treatment and diagnosis of diseases associated
XX with inappropriate secretory protein/membrane protein expression. The
XX nucleic acids and complementary sequences may also be used as DNA probes
XX in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
XX CC and quantitate the presence of similar nucleic acid sequences in samples.
XX CC They may also be used to study the expression and function of secretory
XX CC proteins/membrane polypeptides and their role in metabolism. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC against them and in assays to identify modulators (agonists and
XX CC antagonists) of expression and activity. The antibodies and antagonists
XX CC may also be used as therapeutic agents to down regulate expression and
XX CC activity. The antibodies may also be used as diagnostic agents for
XX CC detecting the presence of the polypeptides in samples (e.g. by enzyme
XX CC linked immunosorbent assay (ELISA). Examples of diseases which may be
XX CC treated include Rheumatoid arthritis and diabetes
XX
XX SQ Sequence 285 AA;
XX
XX Query Match 100.0%; Score 913; DB 4; Length 285;
XX Best Local Similarity 100.0%; Pred. No. 3e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFRRVREFVGVLPPLGTGALRASLVHVGSRPYTEPFPGQH 60
XX |||||
XX DB 32 RGDLMFLDSSASVSHYEFRRVREFVGVLPPLGTGALRASLVHVGSRPYTEPFPGQH 91
XX |||||
XX QY 61 SSGEAAQDAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 120
XX |||||
XX DB 92 SSGEAAQDAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 151
XX |||||
XX QY 121 GPPMDELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDLHIIVOELKGSITLD 180
XX |||||
XX DB 152 GPPMDELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDLHIIVOELKGSITLD 211
XX |||||
XX
XX RESULT 8
XX ID ADY63045 standard; protein; 285 AA.
XX
XX ADY63045;
AC


```
XX 02-JUN-2005 (first entry)
DT
XX Human clone PSEC0053 protein, SEQ ID 48.
DE
XX
XX Gene therapy.
KM
XX Homo sapiens.
OS
XX EPI514933-A1.
PN
XX 16-MAR-2005.
PD
XX 07-JUL-2000; 2004EP-00027228.
XX
XX 08-JUL-1999; 99JP-00194179.
PR 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
PR 07-JUL-2000; 2000EP-00114090.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
XX Ota T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI WPI; 2005-203865/22.
DR N-PSDB; ADY63044.
XX
XX Novel isolated polynucleotide encoding human secretory proteins or
PT membrane proteins, useful for examination and diagnosis of abnormality of
XX human secretory proteins.
XX
XX Disclousure; SEQ ID NO 48; 1240pp; English.
XX
XX The present invention relates to novel human secretory proteins or
XX membrane proteins, and their coding sequences. The present sequence is
XX one such protein sequence. The coding sequences of the invention are
XX useful for examination and diagnosis of abnormality of the human
XX secretory proteins and in gene therapy methods. The coding sequences and
XX proteins are useful as candidates for medicines or as target molecules
XX for developing medicines. Antibodies against the proteins of the
XX invention are useful for treating diseases that are associated with the
XX proteins. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained from sequence information
XX supplied by the European Patent Office.
XX
XX Sequence 285 AA;
SQ
XX
XX Query Match 100.0%; Score 913; DB 9; Length 285;
XX Best Local Similarity 100.0%; Pred. No. 3e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFRRVREFVQGLVAPLPGLGTGALRASLVHVGSRPYTFEPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFRRVREFVQGLVAPLPGLGTGALRASLVHVGSRPYTFEPFGQH 91
XX
XX 61 SSGEAAQDAVRAAQRMGDTHTGTALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVRAAQRMGDTHTGTALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
XX
XX 121 GPPMELKDLGVTYTVIVSTGRGNFLFLSAASAPAEKHLHFVDDDLHITVOELRGSTILD 180
OY 152 GPPMELKDLGVTYTVIVSTGRGNFLFLSAASAPAEKHLHFVDDDLHITVOELRGSTILD 211
DB
XX
XX RESULT 9
XX ID AAE32502 standard; protein; 418 AA.
XX
XX AAE32502;
XX
XX 24-MAR-2003 (first entry)
DT
XX Human Willebrand Factor A domain related-protein (WARP).
```

```
XX Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
KM extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
XX gene therapy; human.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..18
FT /label= Signal_peptide
FT Protein 19..418
FT /note= "Human mature WARP protein"
FT Modified-site 148
FT /note= "O-glycosylation site"
FT Misc-difference 210..211
FT /note= "Encoded by CTCGCG"
FT Modified-site 264
FT /note= "N-glycosylation site"
FT Modified-site 359
FT /note= "N-glycosylation site"
FT Modified-site 361
FT /note= "O-glycosylation site"
FT Disulfide-bond 369..393
XX
XX WO200288184-A1.
XX
XX 07-NOV-2002.
XX
XX 02-MAY-2002; 2002WO-AU000542.
XX
XX 02-MAY-2001; 2001AU-00004701.
XX
XX (MURDOCH CHILDRENS RES INST.
XX
XX Bateman JF, Fitzgerald DJ;
XX WPI; 2003-111873/10.
XX
XX N-PSDB; AAD50399.
XX
XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
PT the manufacture of a medicament in the treatment of a disease condition
PT of the extracellular matrix, in particular arthritis.
XX
XX Claim 11; Page 76-78; 103pp; English.
XX
XX The invention relates to Willebrand Factor A domain related-protein
XX (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
XX superfamily of extracellular matrix (ECM) proteins. WARP is used as a
XX molecular marker, used for detecting a loss of ECM integrity in an animal
XX subject, monitoring repair, regeneration or other disease processes in an
XX animal subject and detecting a disease condition or a propensity for the
XX development of a disease condition in an animal subject. The invention is
XX useful for the manufacture of a medicament in the treatment of a disease
XX condition of the ECM. The disease condition involves the cartilage, and
XX is preferably arthritis. The invention is also used in gene therapy. The
XX present sequence is human WARP protein
XX
XX Sequence 418 AA;
SQ
XX
XX Query Match 100.0%; Score 913; DB 6; Length 418;
XX Best Local Similarity 100.0%; Pred. No. 5.2e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFRRVREFVQGLVAPLPGLGTGALRASLVHVGSRPYTFEPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFRRVREFVQGLVAPLPGLGTGALRASLVHVGSRPYTFEPFGQH 91
XX
XX 61 SSGEAAQDAVRAAQRMGDTHTGTALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVRAAQRMGDTHTGTALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
XX
XX 121 GPPMELKDLGVTYTVIVSTGRGNFLFLSAASAPAEKHLHFVDDDLHITVOELRGSTILD 180
OY
XX
```

DB 152 GPPMOELKDLGTVTFIVSTGRGNFLELSAASAPAEKHLHFVDDHLHIIVQLRGSIID 211

RESULT 10

ABP69674
ID ABP69674 standard; protein; 445 AA.

XX ABP69674;

DT 20-JAN-2003 (first entry)

XX Human polypeptide SEQ ID NO 1721.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
KW antithrptic.

XX Homo sapiens.

XX WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002WO-US005095.

XX 05-MAR-2001; 2001US-00799451.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Zhou P, Goodrich RM, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,
PI Wehrman T, Wang J, Wang D, Dimaenac RT;
XX WPI; 2002-759812/82.
XX N-PSDB; ABZ11891.

XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.

XX Claim 9; SEQ ID NO 1721; 1012pp + Sequence listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences (ABZ11119-
CC ABZ12066) or their mature protein coding portion, active domain coding
CC protein or complementary sequences. The polynucleotides are useful for
CC identifying expressed genes or for physical mapping of human genome. The
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
CC markers, as a food supplement, for generating antibodies, in medical
CC imaging, screening and diagnostic assays and for treating cell-
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 445 AA;

XX Query Match 100.0%; Score 913; DB 5; Length 445;

XX Best local Similarity 100.0%; Pred. No. 5.7e-94;

XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 RGDMLFLDSSASVSHYERSRVEFGQLVAPLPGTGALRASLVHVGSRPYTEPPFGH 60

DB 32 RGDMLFLDSSASVSHYERSRVEFGQLVAPLPGTGALRASLVHVGSRPYTEPPFGH 91

QY 61 SSGEAAQDAVRASAPQMDPTHTGLALVYAKEQLFAEASGARGVPKVLVWTDGSSDPV 120
DB 92 SSGEAAQDAVRASAPQMDPTHTGLALVYAKEQLFAEASGARGVPKVLVWTDGSSDPV 151

QY 121 GPPMOELKDLGTVTFIVSTGRGNFLELSAASAPAEKHLHFVDDHLHIIVQLRGSIID 180
DB 152 GPPMOELKDLGTVTFIVSTGRGNFLELSAASAPAEKHLHFVDDHLHIIVQLRGSIID 211

RESULT 11

ID ADH71106 standard; protein; 445 AA.

XX ADH71106;

DT 25-MAR-2004 (first entry)

XX Human protein of the invention NOVA SEQ ID NO:2.

XX human; cytostatic; immunomodulator; neuroprotective; nootropic;
KW anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;
KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
KW obesity; diabetes; infectious disease; metabolic syndrome X;
KW dyslipidaemia.

XX Homo sapiens.

XX WO2003102155-A2.

XX 11-DEC-2003.

XX 03-JUN-2003; 2003WO-US017430.

XX 03-JUN-2002; 2002US-0385120P.

XX 04-JUN-2002; 2002US-0385784P.

XX 05-JUN-2002; 2002US-0386041P.

XX 06-JUN-2002; 2002US-0386047P.

XX 06-JUN-2002; 2002US-0386453P.

XX 06-JUN-2002; 2002US-0386864P.

XX 06-JUN-2002; 2002US-0387016P.

XX 07-JUN-2002; 2002US-0386796P.

XX 07-JUN-2002; 2002US-0386916P.

XX 07-JUN-2002; 2002US-0386931P.

XX 07-JUN-2002; 2002US-0386942P.

XX 07-JUN-2002; 2002US-0386971P.

XX 08-JUN-2002; 2002US-0296960P.

XX 10-JUN-2002; 2002US-0387400P.

XX 10-JUN-2002; 2002US-0387535P.

XX 11-JUN-2002; 2002US-0387610P.

XX 11-JUN-2002; 2002US-0387625P.

XX 11-JUN-2002; 2002US-0387634P.

XX 11-JUN-2002; 2002US-0387668P.

XX 11-JUN-2002; 2002US-0387702P.

XX 11-JUN-2002; 2002US-0387836P.

XX 11-JUN-2002; 2002US-0387859P.

XX 12-JUN-2002; 2002US-0387933P.

XX 12-JUN-2002; 2002US-0387934P.

XX 12-JUN-2002; 2002US-0387960P.

XX 12-JUN-2002; 2002US-0388022P.

XX 12-JUN-2002; 2002US-0388096P.

XX 13-JUN-2002; 2002US-0389118P.

XX 14-JUN-2002; 2002US-0389120P.

XX 14-JUN-2002; 2002US-0389144P.

XX 14-JUN-2002; 2002US-0389146P.

XX 17-JUN-2002; 2002US-0389729P.

XX 17-JUN-2002; 2002US-0389742P.

XX 18-JUN-2002; 2002US-0389884P.

PR 19-JUN-2002; 2002US-0390066P.
 PR 19-JUN-2002; 2002US-0390209P.
 PR 21-JUN-2002; 2002US-0390763P.
 PR 17-JUL-2002; 2002US-0396706P.
 PR 06-AUG-2002; 2002US-0401628P.
 PR 09-AUG-2002; 2002US-0402155P.
 PR 09-AUG-2002; 2002US-0402256P.
 PR 09-AUG-2002; 2002US-0402389P.
 PR 12-AUG-2002; 2002US-0402816P.
 PR 12-AUG-2002; 2002US-0402821P.
 PR 12-AUG-2002; 2002US-0402833P.
 PR 13-AUG-2002; 2002US-0403448P.
 PR 13-AUG-2002; 2002US-0403459P.
 PR 13-AUG-2002; 2002US-0403531P.
 PR 13-AUG-2002; 2002US-0403532P.
 PR 13-AUG-2002; 2002US-0403563P.
 PR 13-AUG-2002; 2002US-0406317P.
 PR 15-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406355P.
 PR 27-AUG-2002; 2002US-0406240P.
 PR 12-SEP-2002; 2002US-0410084P.
 PR 20-SEP-2002; 2002US-0412528P.
 PR 23-SEP-2002; 2002US-041731P.
 PR 30-SEP-2002; 2002US-041801P.
 PR 30-SEP-2002; 2002US-041839P.
 PR 30-SEP-2002; 2002US-041840P.
 PR 30-SEP-2002; 2002US-041954P.
 PR 09-OCT-2002; 2002US-0417186P.
 PR 09-OCT-2002; 2002US-0417406P.
 PR 23-OCT-2002; 2002US-0420639P.
 PR 28-OCT-2002; 2002US-0421156P.
 PR 31-OCT-2002; 2002US-0422690P.
 PR 01-NOV-2002; 2002US-0423130P.
 PR 05-NOV-2002; 2002US-00437398.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 12-NOV-2002; 2002US-0425453P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Alshbrook JP, Alvarez E, Anderson DM, Boldog FL, Casman SJ;
 PI Catterton E, Chappoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
 PI Eitenberg S, Gangoli EA, Grubich VL, Gorman L, Gunther E, Guo X;
 PI Gusev VY, Herrmann JL, W, Kekuda R, Li L, Liu X, Macdougall JR;
 PI MacLachlan T, Maljanekar UM, Mezick AJ, Miller I, Mishra VS;
 PI Padigaru M, Patrajan M, Pena CE, Peyman JA, Raha D, Rastrelli L;
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
 PI Smithson G, Spytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M;
 PI Zhong H;
 XX
 XX WPI; 2004-081935/08.
 DR N-PSDB; ADH71105.
 XX
 PT New NOVA polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVA-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 PS Example 1; SEQ ID NO 2; 1880pp; English.
 XX
 CC The invention relates to a novel isolated polypeptide (NOVA). A
 CC polypeptide of the invention has cytostatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antipneumatic activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOVA polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,
 CC treating or preventing NOVA-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
 CC further used as hybridization probes, in chromosome mapping, tissue

CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC represents a NOVA polypeptide of the invention.
 XX
 SQ Sequence 445 AA;
 Query Match 100.0%; Score 913; DB 8; Length 445;
 Best Local Similarity 100.0%; Pred. No. 5.7e-94;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RGDLMFLDSSASVSHVEFSRREFGQLVPLPLGICALASLVHVGSRPYTFPPGQH 60
 DB 32 RGDLMFLDSSASVSHVEFSRREFGQLVPLPLGICALASLVHVGSRPYTFPPGQH 91
 QY 61 SSGEAAQDAVVASAQRMGDTHTGALVYAKQLFAEASGAPGVPKVLVWTTDGGSSDPV 120
 DB 92 SSGEAAQDAVVASAQRMGDTHTGALVYAKQLFAEASGAPGVPKVLVWTTDGGSSDPV 151
 QY 121 GPPMOELKDLGVTVFTVSTGNGFLELSAASAPAEKHLHFVDDDLHTIVQELRGSLTD 180
 DB 152 GPPMOELKDLGVTVFTVSTGNGFLELSAASAPAEKHLHFVDDDLHTIVQELRGSLTD 211
 RESULT 12
 AAE03654
 ID AAE03654 standard; protein; 185 AA.
 XX
 AC AAE03654;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Human extracellular matrix and cell adhesion molecule-18 (XMAD-18).
 KW Human; extracellular matrix and cell adhesion molecule; XMAD;
 KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 KW infection; cell proliferative disorder; actinic keratosis; myeloma;
 KW arteriosclerosis; nootropic; anticonvulsant; antithyroid; nephrotoxic;
 KW neuroprotective; dermatological.
 KW
 OS Homo sapiens.
 XX
 FH Homo sapiens.
 FT Key
 FT Region 1..171
 FT /note= "Location/Qualifiers"
 FT Region 1..170
 FT /note= "Collagen glycoprotein precursor"
 FT Region 2..15
 FT /note= "Von Willebrand factor domain score"
 FT Region 37..51
 FT /note= "Collagen glycoprotein precursor"
 FT Region 103..111
 FT /note= "Collagen glycoprotein precursor"
 XX
 PN WO200142285-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 05-DEC-2000; 2000WO-US032990.
 XX
 PR 10-DEC-1999; 99US-0172852P.
 PR 16-DEC-1999; 99US-0172354P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 PI Baughn MR, Lu DM, Shah P, Au-Young J;
 XX

DR MPI, 2001-381632/40.
DR N-PSDB; AAD08062.
XX
XX New human extracellular matrix and cell adhesion molecules and
PT polynucleotide sequences encoding them, useful for diagnosis, prevention,
PT treatment of genetic, autoimmune and cell proliferative disorders.
XX
XX Claim 1; Page 111; 135pp; English.
XX
XX The present sequence is a human extracellular matrix and cell adhesion
CC molecule (XMAP). The XMAP is used for screening a compound for
CC effectiveness as an agonist or antagonist of XMAP. The identified agonist
CC or antagonist are used for treating a disease or condition associated
CC with decreased or increased expression of functional XMAP. The
CC polynucleotides encoding XMAP are useful in somatic or germline gene
CC therapy to correct a genetic deficiency, to express a conditionally
CC lethal gene product and to express a protein which affords protection
CC against intracellular parasites and also for diagnosis of disorders
CC associated with expression of XMAP. They are also used for generating
CC hybridisation probes useful in mapping the naturally occurring genomic
CC sequences and to create knock in humanised animals (pigs) or transgenic
CC animals (mice or rats) to model human diseases. Oligonucleotide or longer
CC fragments derived from the polynucleotide sequences may be used as
CC elements on a microarray. Antibodies which specifically bind XMAP may be
CC used for the diagnosis of disorders associated with the expression of
CC XMAP, or in assays to monitor patients being treated with XMAP. Diseases
CC diagnosed, prevented or treated include genetic disorders such as
CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
CC disease, myotonic dystrophy, sickle cell anaemia, thalassemia,
CC autoimmune/inflammatory disorders such as acquired immune deficiency
CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
CC atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
CC bacterial, fungal, parasitic, protozoal and helminthic infections and
CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
CC and cancer including breast, bladder, bone marrow, brain and uterus
CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma
CC
XX
XX Sequence 185 AA;
SQ

Query Match 97.7%; Score 892; DB 4; Length 185;
Best Local Similarity 100.0%; Pred. No. 3.9e-92;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MFLDSSASVSHSEFRVREPVQOLVAPLPLGALRASLVNHGSPYTFEPFGQSSSE 64
DB 1 MFLDSSASVSHSEFRVREPVQOLVAPLPLGALRASLVNHGSPYTFEPFGQSSSE 60
QY 65 AADAVRASAGNRGDTHTGTALVYAKQFLFAEASGARPGVPKVLVWTTDGGSSDPVGP 124
DB 61 AADAVRASAGNRGDTHTGTALVYAKQFLFAEASGARPGVPKVLVWTTDGGSSDPVGP 120
QY 125 QELKLDGVTVFIVSTGRGNFLLEISAAASAPAEKGLHFVVDVDDLIHTVQELRGSLD 180
DB 121 QELKLDGVTVFIVSTGRGNFLLEISAAASAPAEKGLHFVVDVDDLIHTVQELRGSLD 176

RESULT 13
AAB42581
ID AAB42581 standard; protein; 299 AA.
XX
XX AAB42581;
AC
DT 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF2345 polypeptide sequence SEQ ID NO:4690.
DE
XX Human; open reading frame: ORFX; detection: cytotoxic; hepatotropic;
XX human; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
XX immunostimulant; thrombolytic; coagulant; vasodilator; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW

KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US0008621.
XX
XX 31-MAR-1999; 99US-0127607P.
XX
XX 02-APR-1999; 99US-0127636P.
XX
XX 05-APR-1999; 99US-0127728P.
XX
XX 30-MAR-2000; 2000US-00540763.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Shinketsu RA, Leach M;
XX
XX MPI; 2000-602362/57.
XX
XX N-PSDB; AAC76790.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 3880-3881; 5507pp; English.
XX
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasodilator; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancer, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX Sequence 299 AA;
SQ

Query Match 96.7%; Score 883; DB 3; Length 299;
Best Local Similarity 98.3%; Pred. No. 8e-91;
Matches 175; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 DLMFLDSSASVSHSEFRVREPVQOLVAPLPLGALRASLVNHGSPYTFEPFGQSS 62
DB 37 DLMFLDSSASVSHSEFRVREPVQOLVAPLPLGALRASLVNHGSPYTFEPFGQSS 96
QY 63 GBAADAVRASAGNRGDTHTGTALVYAKQFLFAEASGARPGVPKVLVWTTDGGSSDPVGP 122
DB 97 GBAADAVRASAGNRGDTHTGTALVYAKQFLFAEASGARPGVPKVLVWTTDGGSSDPVGP 156
QY 123 PNOELKLDGVTVFIVSTGRGNFLLEISAAASAPAEKGLHFVVDVDDLIHTVQELRGSLD 180

Db 157 PMOELKDIGTVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLIITVQELRGSILD 214

RESULT 14

AAE32501
ID AAE32503 standard; protein; 180 AA.

XX AAE32503;

XX 24-MAR-2003 (first entry)

XX Mouse von Willebrand Factor A (VA) domain.

XX Willibrand Factor A domain related-protein; von Willebrand Factor A; VA;
KW extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
XX gene therapy; mouse.

XX Mus sp.

XX Key Location/Qualifiers

XX MISC-difference 1..2 /note= "Encoded by GGG"

XX WO200288184-A1.

XX 07-NOV-2002.

XX 02-MAY-2002; 2002WO-AU000542.

XX 02-MAY-2001; 2001AU-00004701.

XX (MURD-) MURDOCH CHILDRENS RES INST.

XX Bateman JF, Fitzgerald DJ;

XX WPI, 2003-111873/10.

XX N-PsDB; AAD50400.

XX New isolated Willibrand Factor A-related protein polypeptide useful for
PT the manufacture of a medicament in the treatment of a disease condition
PT of the extracellular matrix, in particular arthritis.

XX Claim 9; Page 78-79; 103pp; English.

XX The invention relates to Willibrand Factor A domain related-protein
CC (WARP) which is a member of von Willibrand Factor A (VA)-domain protein
CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a
CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medicament in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
CC present sequence is mouse VA domain

XX Sequence 180 AA;

XX Query Match 84.1%; Score 768; DB 6; Length 180;

XX Best Local Similarity 83.3%; Pred. No. 4e-76; Mismatches 16; Indels 0; Gaps 0;

XX Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

XX 1 RGDLMFLDSSASVSHYFERSVRFEVGLVAPLPLGTGLARSLVHVGSRPYTFPPGQH 60

XX 1 QGDLEFLDSSASVSHYFERSVRFEVGLVAPLPLGTGLARSLVHVGSRPYTFPPGQH 60

XX 61 SSGQAIRDAIRVAPOMGDTNTGLALVAKQLFAEAGARBPVKVLVWVTDGSSDPV 120

XX 121 GPMOELKDIGTVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLIITVQELRGSILD 180

XX 121 GPMOELKDIGTVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLIITVQELRGSILD 180

RESULT 15

AAE32501
ID AAE32501 standard; protein; 415 AA.

XX AAE32501;

XX 24-MAR-2003 (first entry)

XX Mouse Willibrand Factor A domain related-protein (WARP).

XX Willibrand Factor A domain related-protein; von Willebrand Factor A; VA;
KW extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
XX gene therapy; mouse.

XX Mus sp.

XX Key Location/Qualifiers

XX Peptide 1..18 /label= Signal_peptide

XX Protein 19..415 /note= "Mouse mature WARP protein"

XX Modified-site 148 /note= "O-glycosylation site"

XX Modified-site 264 /note= "N-glycosylation site"

XX Modified-site 359 /note= "N-glycosylation site"

XX Modified-site 361 /note= "N-glycosylation site"

XX Disulfide-bond 369..393 /note= "O-glycosylation site"

XX Modified-site 400 /note= "O-glycosylation site"

XX WO200288184-A1.

XX 07-NOV-2002.

XX 02-MAY-2002; 2002WO-AU000542.

XX 02-MAY-2001; 2001AU-00004701.

XX (MURD-) MURDOCH CHILDRENS RES INST.

XX Bateman JF, Fitzgerald DJ;

XX WPI, 2003-111873/10.

XX N-PsDB; AAD50022; AAD50398.

XX New isolated Willibrand Factor A-related protein polypeptide useful for
PT the manufacture of a medicament in the treatment of a disease condition
PT of the extracellular matrix, in particular arthritis.

XX Claim 13; Page 74-75; 103pp; English.

XX The invention relates to Willibrand Factor A domain related-protein
CC (WARP) which is a member of von Willibrand Factor A (VA)-domain protein
CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a
CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medicament in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
CC present sequence is mouse WARP protein

XX Sequence 415 AA;

XX Query Match 84.1%; Score 768; DB 6; Length 415;

XX Best Local Similarity 83.3%; Pred. No. 1.3e-77; Mismatches 16; Indels 0; Gaps 0;

XX Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 13:10:29 ; Search time 36.7224 Seconds
(without alignments)
2048.049 Million cell updates/sec

Title: US-10-699-035A-2

Perfect score: 913
Sequence: 1 RQDLMFLDSSASVSHYEFSS.....FVDDVDDHLITVDELKRSIID 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Published Applications_AA_Main:*
2: /cgn2_6/pcodata/1/pubppaa/US07_PUBCOMB.pep:*
3: /cgn2_6/pcodata/1/pubppaa/US08_PUBCOMB.pep:*
4: /cgn2_6/pcodata/1/pubppaa/US09_PUBCOMB.pep:*
5: /cgn2_6/pcodata/1/pubppaa/US10_PUBCOMB.pep:*
6: /cgn2_6/pcodata/1/pubppaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	913	100.0	180	4	US-10-699-035A-2
2	913	100.0	215	3	US-09-789-561-85
3	913	100.0	215	3	US-09-833-245-2096
4	913	100.0	215	5	US-10-883-936-85
5	913	100.0	226	5	US-09-789-561-165
6	913	100.0	226	5	US-10-883-936-165
7	913	100.0	242	3	US-09-789-561-159
8	913	100.0	242	5	US-10-883-936-159
9	913	100.0	418	4	US-10-699-035A-6
10	913	100.0	418	4	US-10-699-035A-20
11	892	97.7	185	4	US-10-149-819-18
12	892	97.7	185	4	US-10-699-035A-8
13	892	97.7	185	4	US-10-699-035A-31
14	892	97.7	185	4	US-10-699-035A-21
15	892	97.7	185	4	US-10-699-035A-21
16	892	97.7	185	4	US-10-699-035A-21
17	892	97.7	185	4	US-10-699-035A-21
18	892	97.7	185	4	US-10-699-035A-21
19	892	97.7	185	4	US-10-699-035A-21
20	892	97.7	185	4	US-10-699-035A-21
21	892	97.7	185	4	US-10-699-035A-21
22	892	97.7	185	4	US-10-699-035A-21
23	892	97.7	185	4	US-10-699-035A-21
24	892	97.7	185	4	US-10-699-035A-21
25	892	97.7	185	4	US-10-699-035A-21
26	892	97.7	185	4	US-10-699-035A-21
27	892	97.7	185	4	US-10-699-035A-21

28	256.5	28.1	755	3	US-09-919-497-57	Sequence 57, Appl
29	256.5	28.1	1780	4	US-10-115-479-10	Sequence 10, Appl
30	254.5	27.9	1207	4	US-10-408-765A-1591	Sequence 1591, Ap
31	254.5	27.9	1297	4	US-10-187-975-102	Sequence 102, App
32	251.5	27.5	776	4	US-10-000-512-8	Sequence 8, Appl
33	251.5	27.5	776	4	US-10-074-566-8	Sequence 176, App
34	251.5	27.5	782	4	US-10-428-275-176	Sequence 1292, Ap
35	251.5	27.5	794	3	US-09-833-245-1292	Sequence 16, Appl
36	251.5	27.5	896	4	US-10-004-378A-16	Sequence 158, App
37	251.5	27.5	896	4	US-10-428-275-158	Sequence 156, App
38	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
39	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
40	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
41	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
42	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
43	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
44	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
45	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl

ALIGNMENTS

```

RESULT 1
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2

Query Match      100.0%; Score 913; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.5e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQDLMFLDSSASVSHYEFSSVFQQLVAPLPTGTALRASLVVSGRPYTFPPGQH 60
    |||||
DB 1 RQDLMFLDSSASVSHYEFSSVFQQLVAPLPTGTALRASLVVSGRPYTFPPGQH 60
    |||||

QY 61 SSGEAAQDAVRASQRMGDTHTGLALVYAKQLFPAASGARGVPEKVLVWTDGSSDPV 120
    |||||
DB 61 SSGEAAQDAVRASQRMGDTHTGLALVYAKQLFPAASGARGVPEKVLVWTDGSSDPV 120
    |||||

QY 121 GPPMGLKQLGTVTVFVSTGRNPLFELSAAPAKKHAFVDDVDDHLITVDELKRSIID 180
    |||||
DB 121 GPPMGLKQLGTVTVFVSTGRNPLFELSAAPAKKHAFVDDVDDHLITVDELKRSIID 180
    |||||

RESULT 2
US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US2002064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043p1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22

```

```
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-85

Query Match          100.0%; Score 913; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.9e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
QY 61 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 180
DB 152 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 211

RESULT 3
US-09-833-245-2096
; Sequence 2096, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PFS46PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO: 2096
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

Query Match          100.0%; Score 913; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.9e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
QY 61 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 180
DB 152 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 211
```

```
DB 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 180
DB 152 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 211

RESULT 4
US-10-883-936-85
; Sequence 85, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; PRIOR FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-85

Query Match          100.0%; Score 913; DB 5; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.9e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
QY 61 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 180
DB 152 GPPMOELKDLCVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDHLIIYQELRGSLTD 211

RESULT 5
US-09-789-561-165
; Sequence 165, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 165
```

LENGTH: 226
TYPE: PRT
ORGANISM: Homo sapiens
US-09-789-561-165

Query Match
Best Local Similarity 100.0%; Score 913; DB 3; Length 226;
Pred. No. 2e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRREFFVGLVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 60
DB 45 RGDLMFLDSSASVSHYEFSSRREFFVGLVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 104
QY 61 SSGEAAQADAVRASACQMGDTHTGLALVYAKQQLFAEASGARPGVPCVLVWVTGSSSDPV 120
DB 105 SSGEAAQADAVRASACQMGDTHTGLALVYAKQQLFAEASGARPGVPCVLVWVTGSSSDPV 164
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 180
DB 165 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 224

RESULT 6

US-10-883-936-165
Sequence 165, Application US/10883936
Publication No. US20050019866A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: PZ043P1
CURRENT APPLICATION NUMBER: US/10/883,936
CURRENT FILING DATE: 2004-07-06
PRIOR APPLICATION NUMBER: US/09/789,561
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 165
LENGTH: 226
TYPE: PRT
ORGANISM: Homo sapiens
US-10-883-936-165

Query Match
Best Local Similarity 100.0%; Score 913; DB 5; Length 226;
Pred. No. 2e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRREFFVGLVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 60
DB 45 RGDLMFLDSSASVSHYEFSSRREFFVGLVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 104
QY 61 SSGEAAQADAVRASACQMGDTHTGLALVYAKQQLFAEASGARPGVPCVLVWVTGSSSDPV 120
DB 105 SSGEAAQADAVRASACQMGDTHTGLALVYAKQQLFAEASGARPGVPCVLVWVTGSSSDPV 164
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 180
DB 165 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 224

RESULT 7

US-09-789-561-159
Sequence 159, Application US/09789561
Patent No. US20020064818A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: PZ043P1

CURRENT APPLICATION NUMBER: US/09/789,561
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 159
LENGTH: 242
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-159

Query Match
Best Local Similarity 100.0%; Score 913; DB 3; Length 242;
Pred. No. 2.2e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRREFFVGLVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 60
DB 59 RGDLMFLDSSASVSHYEFSSRREFFVGLVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 118
QY 61 SSGEAAQADAVRASACQMGDTHTGLALVYAKQQLFAEASGARPGVPCVLVWVTGSSSDPV 120
DB 119 SSGEAAQADAVRASACQMGDTHTGLALVYAKQQLFAEASGARPGVPCVLVWVTGSSSDPV 178
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 180
DB 179 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 238

RESULT 8

US-10-883-936-159
Sequence 159, Application US/10883936
Publication No. US20050019866A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: PZ043P1
CURRENT APPLICATION NUMBER: US/10/883,936
CURRENT FILING DATE: 2004-07-06
PRIOR APPLICATION NUMBER: US/09/789,561
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 159
LENGTH: 242
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE

```
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-159
```

```
Query Match
Best Local Similarity 100.0%; Score 913; DB 5; Length 242;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 RGDLMFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 60
DB 59 RGDLMFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 118
QY 61 SSGEAAQDAVARASQRMGDTHTGLALVYAKQLFAEASGARGVPKVLVWTTDGGSSDPV 120
DB 119 SSGEAAQDAVARASQRMGDTHTGLALVYAKQLFAEASGARGVPKVLVWTTDGGSSDPV 178
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
DB 179 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 238
```

```
RESULT 9
US-10-699-035A-6
Sequence 6, Application US/10699035A
Publication No. US20040214349A1
GENERAL INFORMATION:
```

```
APPLICANT: Bateman, John
APPLICANT: Fitzgerald, David
TITLE OF INVENTION: A Molecular Marker
FILE REFERENCE: A36056 PCT USA A 071838.0142
CURRENT APPLICATION NUMBER: US/10/699,035A
PRIOR FILING DATE: 2003-10-31
PRIOR APPLICATION NUMBER: PCT/AU02/00542
PRIOR FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: AU PR4701/01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 418
TYPE: PRT
ORGANISM: Homo sapiens
US-10-699-035A-6
```

```
Query Match
Best Local Similarity 100.0%; Score 913; DB 4; Length 418;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 RGDLMFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 91
QY 61 SSGEAAQDAVARASQRMGDTHTGLALVYAKQLFAEASGARGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVARASQRMGDTHTGLALVYAKQLFAEASGARGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
DB 152 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211
```

```
RESULT 10
US-10-699-035A-20
Sequence 20, Application US/10699035A
Publication No. US20040214349A1
GENERAL INFORMATION:
```

```
APPLICANT: Bateman, John
APPLICANT: Fitzgerald, David
TITLE OF INVENTION: A Molecular Marker
```

```
FILE REFERENCE: A36056 PCT USA A 071838.0142
CURRENT APPLICATION NUMBER: US/10/699,035A
CURRENT FILING DATE: 2003-10-31
PRIOR APPLICATION NUMBER: PCT/AU02/00542
PRIOR FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: AU PR4701/01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 418
TYPE: PRT
ORGANISM: Homo sapiens
US-10-699-035A-20
```

```
Query Match
Best Local Similarity 100.0%; Score 913; DB 4; Length 418;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 RGDLMFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 91
QY 61 SSGEAAQDAVARASQRMGDTHTGLALVYAKQLFAEASGARGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVARASQRMGDTHTGLALVYAKQLFAEASGARGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
DB 152 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211
```

```
RESULT 11
US-10-149-819-18
```

```
Sequence 18, Application US/10149819
Publication No. US20030044913A1
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: YUE, Henry
APPLICANT: AZIMZAI, Yalda
APPLICANT: TANG, Y. Tom
APPLICANT: PATTERSON, Chandra
APPLICANT: BAUGHN, Mariah R.
APPLICANT: LU, Dyrung Alina M.
APPLICANT: SHAH, Preeti
APPLICANT: IAL, Preeti
APPLICANT: AU-YOUNG, Janice
APPLICANT: BURFORD, Neil
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
FILE REFERENCE: PF-0760 PCT
CURRENT APPLICATION NUMBER: US/10/149,819
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
PRIOR FILING DATE: 1999-12-10; 1999-12-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PERL Program
SEQ ID NO 18
LENGTH: 185
TYPE: PRT
ORGANISM: Homo sapiens
```

```
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CD1
US-10-149-819-18
```

```
Query Match
Best Local Similarity 97.7%; Score 892; DB 4; Length 185;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5 MFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQSSGE 64
DB 1 MFLDSSASVSHYEFSSRVREFVQGVAPLPGLGTALRASLVHVGSRPYTEFPFGQSSGE 60
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:47:59 ; Search time 27.6923 Seconds
(without alignments)
537.392 Million cell updates/sec

Title: US-10-699-035a-2

Perfect score: 913
Sequence: 1 RGDMLFLDSSASVSHREFS.....FVDYDHLHIVQELRGLD 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/6_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/H_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/BCTUS_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/RE_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	384	42.1	77	US-09-513-999C-6277	Sequence 6277, Ap
2	259	28.4	584	US-09-949-016-10340	Sequence 10340, A
3	259	28.4	584	US-09-949-016-10341	Sequence 10341, A
4	256.5	28.1	755	US-09-919-497-57	Sequence 57, Appl
5	251.5	27.5	638	US-08-897-443-1	Sequence 1, Appl
6	251.5	27.5	915	US-09-907-794A-34	Sequence 34, Appl
7	251.5	27.5	915	US-09-905-125A-34	Sequence 34, Appl
8	251.5	27.5	915	US-09-902-775A-34	Sequence 34, Appl
9	251.5	27.5	915	US-09-906-700-34	Sequence 34, Appl
10	251.5	27.5	915	US-09-903-603A-34	Sequence 34, Appl
11	251.5	27.5	915	US-09-904-920A-34	Sequence 34, Appl
12	251.5	27.5	915	US-09-909-064-34	Sequence 34, Appl
13	251.5	27.5	915	US-09-905-381A-34	Sequence 34, Appl
14	251.5	27.5	915	US-09-906-618-34	Sequence 34, Appl
15	251.5	27.5	915	US-09-906-646-34	Sequence 34, Appl
16	251.5	27.5	915	US-09-904-462-34	Sequence 34, Appl
17	251.5	27.5	915	US-09-902-736A-34	Sequence 34, Appl
18	251.5	27.5	915	US-09-906-722A-34	Sequence 34, Appl
19	251.5	27.5	956	US-09-949-016-6215	Sequence 6215, Ap
20	251.5	27.5	963	US-09-949-016-11519	Sequence 11519, A
21	251.5	27.5	963	US-09-949-016-11520	Sequence 11520, A
22	242.5	26.6	956	US-08-897-443-3	Sequence 3, Appl
23	233.5	25.6	496	US-08-463-128-37	Sequence 37, Appl
24	233.5	25.6	496	US-08-463-180-37	Sequence 37, Appl
25	233.5	25.6	496	US-08-001-078A-1	Sequence 1, Appl
26	233.5	25.6	496	US-08-897-443-4	Sequence 4, Appl
27	233.5	25.6	496	US-08-463-218-1	Sequence 1, Appl

28	233.5	25.6	496	2	US-09-949-016-11306	Sequence 11306, A
29	233.5	25.6	496	4	PCT-US94-00253-1	Sequence 1, Appl
30	231.5	25.4	452	2	US-09-914-259-34	Sequence 34, Appl
31	227	24.9	405	2	US-09-312-283C-374	Sequence 374, Appl
32	224.5	24.6	481	2	US-09-914-259-36	Sequence 36, Appl
33	214.5	23.5	486	2	US-09-914-259-35	Sequence 35, Appl
34	214.5	23.5	486	2	US-09-976-594-278	Sequence 278, Appl
35	214.5	23.5	486	2	US-09-949-016-6216	Sequence 6216, Ap
36	214.5	23.5	507	2	US-09-949-016-9878	Sequence 9878, Ap
37	198.5	21.7	171	2	US-09-996-611D-2	Sequence 2, Appl
38	198.5	21.7	954	2	US-09-996-611D-1	Sequence 1, Appl
39	194	21.2	214	2	US-10-061-658-6	Sequence 6, Appl
40	194	21.2	214	2	US-10-061-658-9	Sequence 9, Appl
41	194	21.2	550	2	US-09-907-794A-227	Sequence 227, App
42	194	21.2	550	2	US-09-905-125A-227	Sequence 227, App
43	194	21.2	550	2	US-09-902-775A-227	Sequence 227, App
44	194	21.2	550	2	US-09-906-700-227	Sequence 227, App
45	194	21.2	550	2	US-09-579-288-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-513-999C-6277
Sequence 6277, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duchet, A.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59, US2, REG
CURRENT APPLICATION NUMBER: US/09/513, 999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 6277
LENGTH: 77
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: 66
OTHER INFORMATION: Xaa=Cys or Gly
US-09-513-999C-6277
Query Match
Best Local Similarity 98.7% Pred. No. 4.8e-36;
Matches 76; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 77 MGDHTGALVYAKQLFAEASGARPGVVKLVWVTGSSDPVPPMOELKDLGVTFI 136
DB 1 MGDHTGALVYAKQLFAEASGARPGVVKLVWVTGSSDPVPPMOELKDLGVTFI 60
QY 137 VSTGRNPLELSAASA 153
DB 61 VSTGRNPLELSAASA 77
RESULT 2
US-09-949-016-10340
Sequence 10340, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01107
CURRENT APPLICATION NUMBER: US/09/949,016
```



```

: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10340
: LENGTH: 584
: TYPE: prt
: ORGANISM: Human
: OS-949-016-10340

```

Query Match	28.4%;	Score 259;	DB 2;	Length 584;
Best Local Similarity	37.2%;	Pred. No. 1.5e-20;		
Matches 67;	Conservative 25;	Mismatches 84;	Indels 4;	Gaps 2

QY	3	DLMFLDSSASVSHYEFSSVREFFVQLVAPDLGCGALPASLVHNGSRPYTFEPGQHS	62
Db	348	DVLVLVDSSKSRPNFELVKRFVNNQIVDFLVSEGGTFVLQPSRSRVRTEFPGRVGT	407
QY	63	GEAADAARASQRMGDTHTGLATLYAAKQLFAASGARG---GVPKVLVWVTDGSSSP	119
Db	408	AAEVNQALVAVYEMRGITMTGLAKRMHSHSEAGARPRALNPRVGLVFTDGRSQD	467
QY	120	VGRPMQLKDQGVYTFIVYSTGRNFLELSAASAPAEKILHFV-DYDDLHIIVQELRGI	178
Db	468	ISWAPAAKKEBITVNVAVGVGKAVAEHELEKISEPELHIVASAPDGTMTLLEENRGI	527

```

RESULT 3
US-09-949-016-10341
; Sequence 10341, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231, 498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10341
; LENGTH: 584
; TYPE: prf
; ORGANISM: Human
; US-09-949-016-10341

```

[illegible]

```

RESULT 4
US-09-919-497-57
; Sequence 57, Application US/09919497
; Patent No. 6773863
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,725
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
; LENGTH: 755
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-919-497-57

```

Query Match 28.1%; Score 256.5; DB 2; Length 755;
Best Local Similarity 32.3%; Pred. No. 4.1e-20;
Matches 53; Conservative 38; Mismatches 72; Indels 1; Gaps 1;

[illegible]

Db

OY 60 HSSEADADAVRASAOQMGDTHGLALYAKEQI PAEASGARGGV PKL VWMTDGGSSDP 119

65 YKTEETLLDAIKHISYKCGNTKGAIKVRDPTLFTASGTRGRI PKIVIVITDGRSODD 124

QY 120 VGPFMQELKDGLGVTVFIYSTGRGNFLELSAASAPAEKHLHFD 163
Db 125 VNKISREMLDGYSI FAIGVADADYSELVISIGSPASARHVFFVD 168

RESULT 5
 US-08-897-443-1
 / Sequence 1, Application US/08897443
 / Patent No. 5981263
 / GENERAL INFORMATION:
 / APPLICANT: Hillman, Jennifer L.
 / APPLICANT: Lal, Preeti
 / APPLICANT: Corley, Neil C.
 / APPLICANT: Shah, Purvi
 / APPLICANT: Kaser, Mathew
 / TITLE OF INVENTION: HUMAN MATRILIN-3
 / NUMBER OF SEQUENCES: 4
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Incyte Pharmaceuticals, Inc.
 / STREET: 3174 Porter Drive
 / CITY: Palo Alto
 / STATE: CA
 / COUNTRY: USA
 / ZIP: 94304
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Diskette
 / COMPUTER: IBM Compatible
 / OPERATING SYSTEM: DOS
 / SOFTWARE: FASTSEQ for Windows Version 2.0
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/897,443
 / FILING DATE: Filed Herewith
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER:
 / FILING DATE:
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Billings, Lucy J.
 / REGISTRATION NUMBER: 36,749
 / REFERENCE/DOCKET NUMBER: PF-0348 US
 / TELECOMMUNICATION INFORMATION:
 /

```

PRIORITY APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-794A-34

Query Match 27.5%; Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred.No.2e-19;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2

OY 1 RGDI.MFLDLSASVSHYEFKSVREVGQVAPLPDGTALNLSVHVGSRPTPEFFGQH 60
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 55 RADLVFIIDSSRSYVTHDYAKKEFIYDILQFLDIDGPVTRGGLQYGSIVNKEFSLKTF 114
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 61 SSGEADQAVRASAQRMGDHTGTGLALVYAKQLFAEASGARP---GVPRVLVMTDGGSS 117
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 118 DPVPPRQGEQLKDLGVTFVIVSTGRGNFLELSAASAPAEKHL---HFVVDLHLIYGE 173
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
175 DSVAEVAAKARDTILILFALIGVGVDNFVTLKISGEHPEDHFLVANFSQITLTSVFOK 234
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 7
US-09-905-125A-34
Sequence 34, Application US/0905125A
Patent No. 6664376
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Deenoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertschen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Macher, Jennie P.

```

```

: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/905/125A
: CURRENT FILING DATE: 2001-07-12
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: PCT/US99/28214
: PRIOR FILING DATE: 1999-11-29
: PRIOR APPLICATION NUMBER: PCT/US99/28313
: PRIOR FILING DATE: 1999-11-30
: PRIOR APPLICATION NUMBER: PCT/US99/28564
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 34
: LENGTH: 915
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-905-125A-34

Query Match      27.5%; Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79;

QY 1 RGDMLFDSSASVSHYFSRVRVGGQLVAPLPGTGLRSLVHVSRRPTFFPGH 60
DB 55 RADLVFIIDSSSVVTHDYAKKEFTVDLQPLDGPVTRVGLQYGSYVNNERSLKT 114
QY 61 SSGEAAQDAVAGSAORMGDTHTGLALVYAKEQLFAEASGARP---GVPRVLVWTVDDGSS 117
DB 115 KRKSEVERAVKVRKRLSTGTMTGLAIQYALNIAFSEABGARLRNVRVIMIVTDGRPQ 174
QY 118 DPVCPGQELKDLGVTVFIVSTGRNPLELSAASAPAEKHU---HFVVDVLDLHIIVOE 173
DB 175 DSVAEVAAKARDTGILIFAIIGVGQVDFNTLKSIGSEPHEDHVFIVANSQSLETTLTSVFOK 234

RESULT 8
US-09-902-775A-34
: Sequence 34, Application US/09902775A
```

```

: Patent No. 6686451
: GENERAL INFORMATION:
: APPLICANT: Genentech, Inc.
: APPLICANT: Aeshkenazi, Avi
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Eaton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gertlisen, Mary E.
: APPLICANT: Goddard, A.
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth, J.
: APPLICANT: Kilgavin, Ivar J.
: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/902,775A
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: PCT/US99/28214
: PRIOR FILING DATE: 1999-11-29
: PRIOR APPLICATION NUMBER: PCT/US99/28313
: PRIOR FILING DATE: 1999-11-30
: PRIOR APPLICATION NUMBER: PCT/US99/28564
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 34
: LENGTH: 915
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-902-775A-34
```

```

1  PRIOR FILING DATE: 1999-11-30
2  PRIOR APPLICATION NUMBER: PCT/US99/28564
3  PRIOR FILING DATE: 1999-12-02
4  PRIOR APPLICATION NUMBER: PCT/US99/28565
5  PRIOR FILING DATE: 1999-12-02
6  PRIOR APPLICATION NUMBER: PCT/US99/30095
7  PRIOR FILING DATE: 1999-12-16
8  PRIOR APPLICATION NUMBER: PCT/US99/30911
9  PRIOR FILING DATE: 1999-12-20
10 PRIOR APPLICATION NUMBER: PCT/US99/30999
11 PRIOR FILING DATE: 1999-12-20
12 PRIOR APPLICATION NUMBER: PCT/US00/00219
13 PRIOR FILING DATE: 2000-01-05
14 NUMBER OF SEQ ID NOS: 423
15 SEQ ID NO 34
16 LENGTH: 915
17 TYPE: PRT
18 ORGANISM: Homo sapiens
19 US-09-906-700-34
20
21 Query Match 27.5%; Score 251.5; DB 2; Length 915;
22 Best Local Similarity 32.2%; Pred. No. 2e-19;
23 Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
```

```

; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-603A-34
```

```

Query Match      27.5% Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

Qy 1 RGDLMFLDSSASVSHVEFSRVREVGQVLAPLPFGTGLRSLVHGSRYTEPFPGQH 60
Db 55 RADLVFTIDSSRSVTHYAKKEFIVDLQFLDIGPVTYVGLIQYGSYVKNESLKTf 114
Qy 61 SSGEAAQDAVRASAGRMCDTHGTALVYAKQLFAEASGARP---GVPKVLVWYTDGSS 117
Db 115 KRKSEVERAVKMRHLSTGTMTGLAIQYALNTAFSEAGARPLRENVPRVIMIVTDGRPQ 174
Qy 118 DPVGPMPQELKDLGVTVFIVSTGRGNFLELSAAASAPAEKHL---HFVDVDDLHIIVQ 173
Db 175 DSVAEVAARQDTGLIFATIGVGQVDFNTLKSIGSEPHEDHVLVANSQIETLTGVQK 234
```

```

RESULT 11
US-09-904-920A-34
; Sequence 34, Application US/09904920A
; Patent No. 6806352
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geider, Hanspeter
; APPLICANT: Gerritsen, Maty E.
```

```

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Macher, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paonli, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,920A
```

```

; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-904-920A-34
```

```

Query Match      27.5% Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

Qy 1 RGDLMFLDSSASVSHVEFSRVREVGQVLAPLPFGTGLRSLVHGSRYTEPFPGQH 60
Db 55 RADLVFTIDSSRSVTHYAKKEFIVDLQFLDIGPVTYVGLIQYGSYVKNESLKTf 114
Qy 61 SSGEAAQDAVRASAGRMCDTHGTALVYAKQLFAEASGARP---GVPKVLVWYTDGSS 117
Db 115 KRKSEVERAVKMRHLSTGTMTGLAIQYALNTAFSEAGARPLRENVPRVIMIVTDGRPQ 174
Qy 118 DPVGPMPQELKDLGVTVFIVSTGRGNFLELSAAASAPAEKHL---HFVDVDDLHIIVQ 173
```

Db. 175 DSVAAVAKARDTGLITFAIGVDFNTLKSIGSEPHEDHVLVANSFQIETLTSVFQK 234

RESULT 12

US-09-909-064-34
Sequence 34, Application US/09909064

Patent No. 6818449

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Desnovers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerltzen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gueney, Austin L.

APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kijavlin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-909-064-34

Query Match 27.5%; Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

QY 1 RGDLMFLDSSASVSHYEFSEKRVFVGLVAPLPFGTALRSVLHNGSRPTFRPFQGH 60
Db 55 RADLVFIIDSSRSVTHYAKKEFIVDILFDLGPVTRVGLLQYSTVKNESLKTFF 114
QY 61 SSGEAAODVVRASARQMDTHGLALVYAKEQLFAEASGARP---GVPEVLVWTTDGS 117
Db 115 KRKSEVERAVKMRHLSTGMTGLAIQYALNARSSEAGARPLRENVPRVIVTDGRQ 174
QY 118 DVGPPMOELKDLGTVTFIVSTGRNPLESAASAPAEKHL---HFVVDLHIIYGE 173
Db 175 DSVAAVAKARDTGLITFAIGVDFNTLKSIGSEPHEDHVLVANSFQIETLTSVFQK 234

RESULT 13

US-09-905-381A-34

Sequence 34, Application US/09905381A

Patent No. 6818746

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Desnovers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerltzen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gueney, Austin L.

APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kijavlin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

APPLICANT: Wood, William, I.

```

; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-381A-34
```

```

Query Match      27.5% Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79;

QY      1  RGDLMFLDSSASVSHYEFSSVRREFVGQVLAFLPLGTGALRASLVHVGSRPYTEFPFGQH 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      55  RADLVFIIDSSRSVTHYDAKVEFIVDILQFLDIGPVTVGLQYGSIVKNESSLKTF 114
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY      61  SSGEAAQDAVNASAQRMGDTHTGLALVYAKQDLPAASGARP---GVPKVLVWVTDDGSS 117
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      115  KRKSEVERAVKRMRLSTGTMTGLAIQYALNIAFSEAGARPLRENVPRVIMIVTDGRQ 174
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY      118  DVGPPMOELKDGLGTVTVIVSTGRGNFLELSAASAPAEKHL---HFVVDLHLIYQE 173
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      175  DSVAEVAAKARDTGILIFALIGVGQVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVQK 234
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

RESULT 14
US-09-906-618-34
; Sequence 34, Application US/09906618
; Patent No. 6828146
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
```

```

; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,618
; CURRENT FILING DATE: 2001-07-15
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-906-618-34
```

```

Query Match      27.5% Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79;
```

```

QY      1  RGDLMFLDSSASVSHYEFSSVRREFVGQVLAFLPLGTGALRASLVHVGSRPYTEFPFGQH 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      55  RADLVFIIDSSRSVTHYDAKVEFIVDILQFLDIGPVTVGLQYGSIVKNESSLKTF 114
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY      61  SSGEAAQDAVNASAQRMGDTHTGLALVYAKQDLPAASGARP---GVPKVLVWVTDDGSS 117
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      115  KRKSEVERAVKRMRLSTGTMTGLAIQYALNIAFSEAGARPLRENVPRVIMIVTDGRQ 174
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY      118  DVGPPMOELKDGLGTVTVIVSTGRGNFLELSAASAPAEKHL---HFVVDLHLIYQE 173
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      175  DSVAEVAAKARDTGILIFALIGVGQVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVQK 234
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

RESULT 15
US-09-906-646-34
; Sequence 34, Application US/09906646
; Patent No. 6852848
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
```



```

1  APPLICANT: Eaton, Dan L.
2  APPLICANT: Ferrara, Napoleone
3  APPLICANT: Filvaroff, Ellen
4  APPLICANT: Fong, Sherman
5  APPLICANT: Gao, Wei-Qiang
6  APPLICANT: Gerber, Hanspeter
7  APPLICANT: Gerritsen, Mary E.
8  APPLICANT: Goddard, A.
9  APPLICANT: Godowski, Paul J.
10 APPLICANT: Grimaldi, Christopher J.
11 APPLICANT: Gurney, Austin L.
12 APPLICANT: Hillan, Kenneth J.
13 APPLICANT: Kijavini, Ivar J.
14 APPLICANT: Mather, Jennie P.
15 APPLICANT: Pan, James
16 APPLICANT: Paoni, Nicholas F.
17 APPLICANT: Roy, Margaret Ann
18 APPLICANT: Stewart, Timothy A.
19 APPLICANT: Tumas, Daniel
20 APPLICANT: Williams, P. Mickey
21 APPLICANT: Wood, William, I.
22 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
23 TITLE OF INVENTION: Acids Encoding the Same
24 FILE REFERENCE: 10466-14
25 CURRENT APPLICATION NUMBER: US/09/906,646
26 CURRENT FILING DATE: 2002-01-22
27 PRIOR APPLICATION NUMBER: PCT/US00/04414
28 PRIOR FILING DATE: 2000-02-22
29 PRIOR APPLICATION NUMBER: US 60/143,048
30 PRIOR FILING DATE: 1999-07-07
31 PRIOR APPLICATION NUMBER: US 60/145,698
32 PRIOR FILING DATE: 1999-07-26
33 PRIOR APPLICATION NUMBER: US 60/146,222
34 PRIOR FILING DATE: 1999-07-28
35 PRIOR APPLICATION NUMBER: PCT/US99/20594
36 PRIOR FILING DATE: 1999-09-08
37 PRIOR APPLICATION NUMBER: PCT/US99/20944
38 PRIOR FILING DATE: 1999-09-13
39 PRIOR APPLICATION NUMBER: PCT/US99/21090
40 PRIOR FILING DATE: 1999-09-15
41 PRIOR APPLICATION NUMBER: PCT/US99/21547
42 PRIOR FILING DATE: 1999-09-15
43 PRIOR APPLICATION NUMBER: PCT/US99/23089
44 PRIOR FILING DATE: 1999-10-05
45 PRIOR APPLICATION NUMBER: PCT/US99/28214
46 PRIOR FILING DATE: 1999-11-29
47 PRIOR APPLICATION NUMBER: PCT/US99/28313
48 PRIOR FILING DATE: 1999-11-30
49 PRIOR APPLICATION NUMBER: PCT/US99/28564
50 PRIOR FILING DATE: 1999-12-02
51 PRIOR APPLICATION NUMBER: PCT/US99/28565
52 PRIOR FILING DATE: 1999-12-02
53 PRIOR APPLICATION NUMBER: PCT/US99/30095
54 PRIOR FILING DATE: 1999-12-16
55 PRIOR APPLICATION NUMBER: PCT/US99/30911
56 PRIOR FILING DATE: 1999-12-20
57 PRIOR APPLICATION NUMBER: PCT/US99/30999
58 PRIOR FILING DATE: 1999-12-20
59 PRIOR APPLICATION NUMBER: PCT/US00/00219
60 PRIOR FILING DATE: 2000-01-05
61 NUMBER OF SEQ ID NOS: 423
62 SEQ ID NO 34
63 LENGTH: 915
64 TYPE: PRT
65 ORGANISM: Homo sapiens
66 US-09-906-646-34

```

Query Match	27.5%	Score 251.5	DB 2	Length 915
Best Local Similarity	32.2%	Pred. No. 2e-19		
Matches 58	Conservative 36	Mismatches 79	Indels 7	Gaps 2

QY 1 RGDMLLDSSASVSHYEFSSRVAEFVGQVLAPLPDGTGALRASLVHGVSSPYTEPFQGH 60

```

Db      55 RADLVPIIDSSSYNTHDYAKXKEITVDILQFLDIGDVTRVGLQGSYKXNEFSLKT 114
OY      61 SSGEAAQDAVRASAOQMGGDTHTGLALVTYAKQLFPAEASGARP---GYPKVLWVTTDGGSS 117
Db      115 KRKSEVERAVKRMRLSTGTMTGLAIQVALINIAFSEAGARPLRENPYIVIMITVDPQP 174
OY      118 DPYGPMPQELKDLGVTVFIVTSGRGNFLELSAASAPAEKGL-----HFVVDVDDHIIVOE 173
Db      175 DSYAEVAARAKRDPGILIFAIIGVGQVDFTLKISIGSEPHEDHVFVLPANFSQIETLTYSVFOK 234

```

Search completed: February 13, 2006, 07:50:51
Job time : 28.6923 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:40:09 ; Search time 62.2107 Seconds
(without alignments)
646,490 Million cell updates/sec

Title: US-10-699-035a-6

Perfect score: 2154

Sequence: 1 MPMWTALGALSRLALARS.....RRPRFPVPRAPPTGRASREP 418

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : PIR 80:*

1: pirl:*

2: pirl:*

3: pirl:*

4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437.5	20.3	3124	2 A40020	collagen alpha 1(X)
2	382.5	17.8	2944	2 A54849	collagen alpha 1(V)
3	380.5	17.7	1857	2 S31212	collagen alpha 1(X)
4	380.5	17.7	1888	2 S78476	collagen alpha 1(X)
5	369.5	17.2	1747	2 A45974	collagen alpha 1(X)
6	278	12.9	929	2 I51027	type XII collagen
7	268.5	12.5	843	2 A40970	undulin 1 - human
8	248.5	11.5	493	2 A33809	cartilage matrix P
9	243.5	11.3	500	2 S66522	cartilage matrix P
10	239	11.1	3137	2 A37797	collagen alpha 3(V)
11	233.5	10.8	496	2 A37979	cartilage matrix P
12	225.5	10.5	3176	2 CGH3A	collagen alpha 3(V)
13	220	10.2	741	2 T46488	hypothetical prote
14	209	9.7	272	2 A55348	integrin alpha-1
15	208	9.7	1151	2 A45226	integrin alpha-1
16	202	9.4	1180	2 A35854	integrin alpha-1
17	183	8.5	1286	2 A48396	protein M01E10.2
18	181.5	8.4	724	2 A48569	antigen Em100 - E1
19	179	8.3	550	2 T23760	hypothetical prote
20	174.5	8.1	2481	2 A43908	fibronectin - Afri
21	174	8.1	1153	1 RWHUB	cell surface glyco
22	173	8.0	1153	1 S00551	leukocyte surface
23	170.5	7.9	1810	1 A32230	tenascin precursor
24	169.5	7.9	3051	1 S42373	hypothetical prote
25	168.5	7.8	1163	1 RWHUB	cell surface glyco
26	168	7.8	712	1 A45618	immunoglobulin mic
27	167	7.8	427	2 G00039	von Willebrand fac
28	166.5	7.7	1746	1 S19694	tenascin precursor
29	165.5	7.7	2201	2 A32160	tenascin-C - human

30	165	7.7	1163	2 I56126	lymphocyte function
31	164	7.6	3566	1 A40701	tenascin-X precurs
32	163.5	7.6	1170	2 I45914	integrin alpha 2 s
33	163	7.6	4135	2 T42629	tenascin-X - bovin
34	162.5	7.5	933	2 A31930	cytotactin - chick
35	162.5	7.5	2386	1 FNHU	fibronectin precur
36	162	7.5	2813	1 VWHU	von Willebrand fac
37	161	7.5	1178	2 S44142	VLA-2 protein homo
38	160	7.4	1907	2 S50893	protein-tyrosine-P
39	157.5	7.3	1020	2 A29355	fibronectin - chic
40	154	7.1	2265	1 FNBO	fibronectin - bovi
41	153.5	7.1	2019	1 J03322	tenascin precursor
42	152	7.1	2477	2 S14428	fibronectin precur
43	151.5	7.0	1807	2 UC6319	integrin beta-4 ch
44	151	7.0	1914	2 T42635	tenascin Y precurs
45	150.5	7.0	1170	2 S03308	cell surface glyco

ALIGNMENTS

RESULT 1

A40020
collagen alpha 1(XII) chain precursor - chicken
N/Alternate names: fibrochimerin
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #ext change 09-Jul-2004
C/Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
R/Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Oba
U. Cell Biol. 115, 209-221, 1991
A/Title: The complete primary structure of type XII collagen shows a chimeric molecule
nous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site
A/Reference number: A40020; M01D:92011862; PMID:1918137
A/Accession: A40020
A/Molecule type: mRNA
A/Residues: 1-3124 <YAM>
A/Cross-references: UNIPROT:P13944; UNIPARC:UPI0000136D2D; GB:D00824; NID:G222810; PIDN
A/Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and
R/Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989
A/Title: Type XII collagen: A large multidomain molecule with partial homology to type
A/Reference number: A34485; M01D:90062079; PMID:2584192
A/Accession: A34485
A/Molecule type: mRNA
A/Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A/Cross-references: UNIPARC:UPI0000171233; EMBL:J05137; NID:G211284; PIDN:AAA48635.1; P
A/Accession: B34485
A/Molecule type: protein
A/Residues: 2772-2792, 2846-2873 <GOR2>
A/Cross-references: UNIPARC:UPI0000173C3F; UNIPARC:UPI0000173C40
R/Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A/Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA
A/Reference number: A28037; M01D:87317590; PMID:3476925
A/Accession: A28037
A/Molecule type: mRNA
A/Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
A/Cross-references: UNIPARC:UPI0000171294; EMBL:M17375; NID:G211649; PIDN:AAA48718.1; P
A/Note: This sequence has been revised in reference A34485
R/Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A/Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A/Reference number: S23814; M01D:92362621; PMID:1323460
A/Accession: S23814
A/Molecule type: protein
A/Residues: 'X', 1335, 'Q', 1335-1347, 1914-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517
A/Cross-references: UNIPARC:UPI0000173C41; UNIPARC:UPI0000173C42; UNIPARC:UPI0000173C43
R/Dublet, B.; van der Rest, M.
J. Biol. Chem. 262, 17724-17727, 1987
A/Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of peptin
A/Reference number: S22254; M01D:88087065; PMID:3121603
A/Accession: S22254
A/Molecule type: protein

A.Residues: 2831-2832,'T',2834,'R',2836-2843,3002-3014 <DUB>
A.Cross-references: UNIPARC:UPI0000173C44; UNIPARC:UPI0000173C45
R:trueb, J.; Trueb, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A.Title: The two splice variants of collagen XII share a common 5' end.
A.Reference number: S28811; MUID:93042014; PMID:1420368
A.Accession: S28811
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-24,1189-1257,'S',1259-1263,'E',1265-1280 <TRU>
A.Cross-references: UNIPARC:UPI0000173C46; EMBL:X67327
C.Genetics:
A.Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
C.Keywords: alternative splicing; cell binding; coiled coil; connective tissue, disulfid
F.1-23/Domain: signal sequence #status predicted <Sig>
F.24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F.24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted
F.24-114/Domain: IIIA #status predicted <IIIA>
F.124-105/Domain: fibronectin type III repeat homology <FN3A>
F.137-301/Domain: von Willebrand factor type A repeat homology <VWA1>
F.332-423/Domain: IIIB #status predicted <IIIB>
F.332-414/Domain: fibronectin type III repeat homology <FN3B>
F.437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
F.629-1178/Domain: IIIC #status predicted <IIIC>
F.630-711/Domain: fibronectin type III repeat homology <FN3C>
F.721-802/Domain: fibronectin type III repeat homology <FN3D>
F.812-895/Domain: fibronectin type III repeat homology <FN3E>
F.905-986/Domain: fibronectin type III repeat homology <FN3F>
F.995-1076/Domain: fibronectin type III repeat homology <FN3G>
F.1086-1169/Domain: fibronectin type III repeat homology <FN3H>
F.1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>
F.1384-1465/Domain: IIID #status predicted <IIID>
F.1474-1557/Domain: fibronectin type III repeat homology <FN3I>
F.1566-1647/Domain: fibronectin type III repeat homology <FN3J>
F.1566-1738/Domain: fibronectin type III repeat homology <FN3K>
F.1755-1838/Domain: fibronectin type III repeat homology <FN3M>
F.1847-1928/Domain: fibronectin type III repeat homology <FN3N>
F.1937-2019/Domain: fibronectin type III repeat homology <FN3O>
F.2028-2110/Domain: fibronectin type III repeat homology <FN3P>
F.2119-2199/Domain: fibronectin type III repeat homology <FN3Q>
F.2207-2294/Domain: fibronectin type III repeat homology <FN3R>
F.2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>
F.2438-2440/Domain: cell adhesion #status predicted
F.2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted
F.2899-2902/Domain: collagenous COL2 #status predicted <COL2>
F.2899-2901/Region: cell attachment (R-G-D) motif
F.2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>
F.2946-3048/Domain: collagenous COL1 #status predicted <COL1>
F.3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>
F.321,1006,1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (asn) (cova
F.2780,2789,2836,2842,2860,2866,2869,3004,3007/Modified site: hydroxyproline (Pro) #stat

Query Match 20.3%; Score 437.5; DB 2; Length 3124;
Best Local Similarity 31.1%; Pred. No. 1.6e-22;
Matches 123; Conservative 66; Mismatches 168; Indels 39; Gaps 8;
QY 32 RGLMPLDLSASVSHYEFGRVPEFGOLVAPLPLCTGMLRSLVGVSGRPYEPFGOH 91
DB 437 KADVFLVGVSGVSIIGIANVYKARFLEVLVKSFIEIPKRVQISLVGYSDPHMFELNXY 496
QY 92 SSGEAAQDAVRASAGQMGDTHTGLALVYAEQLFABAAGARPQVPRVLVWVTDGSSDPV 151
DB 497 NRVKDIQIAINFPYRGSGTNTSKAMTYREKVFVTSKSGRPVPRVMITITDGKSSDAF 556
QY 152 GPPMGLKDLGVTVFVYVSGRGFLELSAASAPARKHLHFV-DVVDLHIYVELAGSL- 209
DB 557 KSPATKLRADVAIFVGVGVDAVRTLELAISPPATHHYVDFDFAFORISPELLQSYVC 616
QY 210 -----LDMNR-----POOLHTEITSSGFRFLAMPPLLTADGGYVLELTPAAGCAAR 257
DB 617 LAIEQLAAIRKKSYPYPAKMVMSVDTTSDSFKYSMSAAGSEKSYLITKYV-AIGDGF 674

QY 238 RQQLPGNATDWMAGLDPTDYVALVPEASNRLRLPQLLRKTRPREAGPERIYSHAR 317
DB 675 IVSVASPTSSVLTNLDELTYVAASVIAE-----YEDGDPPLDGETTLEVXGA 724
QY 318 PRLSR-----VSNAAPALGSAALGYHVFGLRGAEQRYEVPAGRCCTTLOGLA 367
DB 725 PNNLRITBETTDTSFVGGWTPAPGN--VLRRLVYRPLTGGERKQYTVSANENSTTLRLNI 782
QY 368 PGTAYLVYVTAAFNRSGRESALSAKACTPDGPRPRR 403
DB 783 PDRREVSIVAEYQSGPMALNGVAKT-DEVAGNDR 817
RESULT 2
A54849
collagen alpha 1(VII) chain precursor - human
N.Alternate names: procollagen alpha 1(VII) chain
C.Species: Homo sapiens (man)
C.Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #text change 09-Jul-2004
C.Accession: A54849; PH0844; S16316; 156328; A30296; I84686
R.Christiano, A.M.; Greenspan, D.S.; Lee, S.; Ulitto, J.
J. Biol. Chem. 269, 20256-20262, 1994
A.Title: Cloning of human type VII collagen. Complete primary sequence of the alpha(VII
A.Reference number: A54849; MUID:94327588; PMID:8051117
A.Accession: A54849
A.Status: not compared with conceptual translation
A.Molecule type: mRNA
A.Cross-references: UNIPARC:UPI000017A138; GB:L02870; NID:9897124; PIDN:
R.Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A.Title: Molecular cloning and characterization of type VII collagen cDNA.
A.Reference number: PH0844; MUID:92231902; PMID:1567409
A.Accession: PH0844
A.Molecule type: mRNA
A.Residues: 'EPR',340-475,'RALSTASHTLCWRATRMHPCNRGSHWTRAPCEPCNRPASHRAARAG',524-528,'C',
A.Cross-references: UNIPARC:UPI000017A139; DBJ:D11152; DDBJ:D13654; NID:9453658; PIDN:B
A.Experimental source: keratinocyte
A>Note: the authors translated the codon ACC for residues 394 and 397 as Tyr
R.Parente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A.Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A.Reference number: S16316; MUID:91334380; PMID:1871109
A.Accession: S16316
A.Molecule type: mRNA
A.Residues: 815-892,'E',894-1439 <PAR>
A.Cross-references: UNIPARC:UPI000016A722; GB:M65158; GB:S49017; NID:G180914; PIDN:AAA96
A.Experimental source: keratinocyte
R.Gadmon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prineas, P.S.; Cook, M.E.; Wright, J.,
J. Invest. Dermatol. 99, 691-696, 1992
A.Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot
A.Reference number: I56328; MUID:93107742; PMID:1469284
A.Accession: I56328
A>Status: translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 'EPR',372-517,'DV',520-540,'W',542-1255 <RES>
A.Cross-references: UNIPARC:UPI000016B3AC; GB:S51236; NID:9262308; PIDN:AA24637.1; PID
R.Selzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morrill, N.P.; Glatville, R.W.; Burgeson, R.E
J. Biol. Chem. 264, 3822-3826, 1989
A>Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase
A.Reference number: A30296; MUID:89139437; PMID:2537292
A.Accession: A30296
A.Molecule type: protein
A.Residues: 'A',1240-1246,'G',1248-1250,'XE',1253-1255,'Q',1257,'E',2032,'C',2034-2041,'
A.Cross-references: UNIPARC:UPI000017A13A; UNIPARC:UPI000017A13B; UNIPARC:UPI000017A13C;
A>Note: two reported peptides cannot be reliably located
R.Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A>Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A.Reference number: I48103; MUID:93271985; PMID:8499916
A.Accession: I48103
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA

A;Residues: 2395-2871,'S',2873-2944 <RE2>
A;Cross-references: UNIPARC:UPI000016A724; GB:L06862; NID:9388713; PIDN:AAA89196.1; PID:
R:Christiano, A.M.; Ruynaenen, M.; Uitto, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A;Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --- Ser substitution
A;Reference number: A5555; MID:94224777; PMID:8170945
A;Contents: Annotation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GJB:COL7A1; EBR1; EBD1; EB
A;Cross-references: GDB:128750; OMIM:120120
A;Map position: 3p21.3-3p21.3
A;Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis
A;Note: there are 118 introns
C;Complex: type VII collagen is probably a homotrimer
C;Function:
A;Description: Structural component of extracellular polymer associated with anchoring
F;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F;17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F;36-201/Domain: von Willebrand factor type A repeat homology <VMA1>
F;331-318/Domain: fibronectin type III repeat homology <FN1>
F;327-413/Domain: fibronectin type III repeat homology <FN2>
F;314-502/Domain: fibronectin type III repeat homology <FN3>
F;508-593/Domain: fibronectin type III repeat homology <FN4>
F;598-683/Domain: fibronectin type III repeat homology <FN5>
F;686-771/Domain: fibronectin type III repeat homology <FN6>
F;776-862/Domain: fibronectin type III repeat homology <FN7>
F;864-952/Domain: fibronectin type III repeat homology <FN8>
F;954-1045/Domain: fibronectin type III repeat homology <FN9>
F;1052-1219/Domain: von Willebrand factor type A repeat homology <VMA2>
F;1170-1172/Region: cell attachment (R-G-D) motif
F;1189-1253/Region: cysteine/proline-rich
F;1234-2783/Region: interrupted helical
F;1334-1336/Region: cell attachment (R-G-D) motif
F;2008-2010/Region: cell attachment (R-G-D) motif
F;2553-2555/Region: cell attachment (R-G-D) motif
F;2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F;2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F;337,786,1109/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F;2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F;2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;2634,2802,2804/Disulfide bonds: interchain #status predicted

Query Match 17.8%; Score 382.5; DB 2; Length 2944;
Best Local Similarity 31.0%; Pred. No. 1.2e-18;
Matches 136; Conservative 46; Mismatches 208; Indels 49; Gaps 11;

11 LSLRLALRSGARGPASAPR-----GDMFLDSSASVSHYPSRRV 54
1 MTRRLVAALCA--GILAEPRVPAQHREVVCTRLVADIVFLDSSSSIGSNFEVR 58
55 EFVGVAVAPLPLGTGA--LRASLVHVGSRPYTEFPFGQSSGEAQAQVAPASQRMGDT 112
59 SFLEGLVLPFGSAGAAQGVAFATVQYSDPRTEFGDALGSGGVIRALIELSKGNT 118
113 TGLALVYAKQLFAEASGARPGVVKLVWYTDGSSDPVPPMOELKDLGVTVFYVSTGR 172
119 TGAAILHVDHVFPLPOL-ARPGVAKVCILITDGKSQDLVDTAAQLKGGQGVKFAVGIK 177
173 GNLELESAASAPAEKILHVDVDDLHII-----VQELRGSILDMMP 215
178 ADBELKRVASQPTSDFFPF--VNDPSILRTLPLVSRVCTTAGGVVTRPPDDTSAP 235
216 QQLHATEITSSGFRAMPPLITADSGYV--LELVPSAQGAARQO--LPGNATWITWA 271
236 RDLVLSPPSSQSLRVQWTAASGPVTKVQYPLTGLGQPLPBERGVNVPAGETSVRLR 295
272 GLDPTDIDVAVLPESNVRLRPOILKVRTRPEBAGERIVISHARPRLSVSWAPALGS 331

Db 296 GLRPLEYQVTVIALYANSIGENVGSTATTALF-GPE-LITONTAHSILVWRSVPG- 352
Qy 332 AALGYHVOFGPLRGEAQRVEVPAGRNCTTGLAPGATVYLVTAARSGRESALSAR 391
Db 353 --ATGRTVWVRVLSGGPPTQQLGPGQSVLLRLDLRPGTDYETVTVTLGORSVGPAATSLM 410
Qy 392 ACTPDGPRPRPPVPAAPT 410
Db 411 ARTDASQVETLRPVILGPT 429

RESULT 3
S31212
collagen alpha 1(XIV) chain precursor, short form - chicken
C;Species: Gallus gallus (chicken)
C;Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #ext_change 15-Sep-2003
C;Accession: S31212
R;Weilnhi, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A;Title: Complete primary structure of chicken collagen XIV.
A;Reference number: S31211; MID:93185668; PMID:8444186
A;Accession: S31212
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-1857 <MA>
A;Cross-references: UNIPARC:UPI000017123C; EMBL:X70792; NID:9288874; PIDN:CAA50063.1; P
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Genes: COL14A1
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>
F;29-110/Domain: fibronectin type III repeat homology <FN3A>
F;156-320/Domain: von Willebrand factor type A repeat homology <VMA1>
F;352-433/Domain: fibronectin type III repeat homology <FN3B>
F;442-525/Domain: fibronectin type III repeat homology <FN3C>
F;534-614/Domain: fibronectin type III repeat homology <FN3D>
F;623-707/Domain: fibronectin type III repeat homology <FN3E>
F;741-823/Domain: fibronectin type III repeat homology <FN3F>
F;832-914/Domain: fibronectin type III repeat homology <FN3G>
F;922-1009/Domain: fibronectin type III repeat homology <FN3H>
F;1040-1205/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 17.7%; Score 380.5; DB 2; Length 1857;
Best Local Similarity 28.6%; Pred. No. 9.1e-19;
Matches 117; Conservative 63; Mismatches 158; Indels 71; Gaps 13;

27 PASAPRGLMFLDSSASVSHYPSRRVREFVGVAPLPLGTALASLVHVGSRPYTER 86
154 PAIA--DIVILVDGSMISGRFNLVRLFLENLVAFNVGSEKTVGLAQVSGDPRIEW 210
211 HLNAVGTAKVLDVAVNMLPYKGSNTLTGLALTYILENSFKPEAGAPGSKILITTDK 270
147 SSDPVPMPMOELKDLGVTVFYVSTGRNLELSAASAPAEKILAFV-DVDDLHIIVQ-- 203
271 SQDDVLPKAKMLRDAIBFALGVKNADINELKEIASBPSTHVWVADPNFMSIVEGL 330
204 -----ELRGSILDM-RPQQLHATEITSSGFRAMPPLITADSG----- 241
331 TRTVCSRVBEKEIKGTIAASLGAPTDLVTSITARGRVSV-----THSPGVEXKRVV 386
242 YVVLLELVPSAQAARQOOLPGNATWITWAGLDPDDIDVAVL-----PESN 288
387 YI-----PTRGGQPEVAVVDDSSSTAVLKNLMSLIEYQIAYVAITSNAASEGLRGETT 440
289 VRLRLRPOILKVRTRPEBAGERIVISHARPRLSVWAPALGSAAALGYHVOFGPLRG- 347
441 LAL-----PMAADLKLYDVSHS---SWRAKMN---GVAGATGVWILYALPLDEGL 483
348 EAQRVEVPAGRNCTTQ--GLAPGATVYLVTAARSGRESALSARACT 394

Db 484 MADEKEIKIGEAStELDLGLPNTYTVYVYAMFGEASDPLTGQETT 532

RESULT 4

S78476
collagen alpha 1(XIV) chain precursor, long form - chicken
C/Species: Gallus gallus (chicken)
C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
C/Accession: S78476; S31211
J. Trueb, B.
submitted to the EMBL Data Library, January 1993
A/Reference number: S78476
A/Accession: S78476
A/Molecule type: mRNA
A/Residues: 1-1888 <TRU>
A/Cross-references: UNIPROT:P32018; UNIPARC:UPI0000126D31; EMBL:X70793; NID:G288872; PID:R.Maechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A/Title: Complete primary structure of chicken collagen XIV.
A/Reference number: S31211; MUID:93185668; PMID:8444186
A/Accession: S31211
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-416;1460-1811,1843-1888 <MAE>
A/Cross-references: UNIPARC:UPI00001773F4; UNIPARC:UPI00001773F5; EMBL:X70793
C/Genes: COL14A1

C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F.1-28/Domain: signal sequence #status predicted <Sig>
F.29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <Mat>
F.29-110/Domain: fibronectin type III repeat homology <FN3A>
F.156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F.352-433/Domain: fibronectin type III repeat homology <FN3C>
F.442-525/Domain: fibronectin type III repeat homology <FN3C>
F.534-614/Domain: fibronectin type III repeat homology <FN3D>
F.623-707/Domain: fibronectin type III repeat homology <FN3D>
F.741-823/Domain: fibronectin type III repeat homology <FN3D>
F.832-914/Domain: fibronectin type III repeat homology <FN3D>
F.922-1009/Domain: fibronectin type III repeat homology <FN3D>
F.1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 17.7%; Score 380.5; DB 2; Length 1888;
Best Local Similarity 28.6%; Pred. No. 9.3e-19;
Matches 117; Conservative 63; Mismatches 158; Indels 71; Gaps 13;

27 PASAPRGDMLPDLSSASVSHYFRRFVGQVLPPLGTGALRASLVHGSRRPYTF 86
154 PAIA---DIVILVDGWSIGRNFRLVRLFLNLVSAFVSGSKTRVGLAQSGDPRIM 210
87 PFGQSSGEAODAVRASQRMGDHTGTALVYAKQFLPAEASGARPGVPKYLWVTDG 146
211 HLNAGYTKDAVLDVAVNLPRYKGGNTLTGLALTYLENSFKPEAGAPGVSKIGILTDDK 270
147 SSDPVGPPMOELKDLGVTFYVSTGRNFLELSAASAAPAEKHLHV-DVDDLHIVQ-- 203
271 SDDDVTPPAKNLRDAIEFLAIGVKNADINEKEIASSEDSHTVYVAVDFNFNSIVEGL 330
204 -----ELRGSILDM-PPQQLHATEITSSGRLAMPPLTLDG----- 241
331 TRTVGSRVEEQKEIKGTIAASIGAPTDLVTSITARGRVN-----TSPKVEKRYRV 386
242 YVLELVPSAQPGAAARQQLPGNATDWIAGLDPTDIDVALV-----PESN 288
387 YV-----PTRGGQPREVVVDGSSPAVKNLMSLTREYQIAVFAISNAASEGLRGTEET 440
289 VALLRPQILIRKTRPEAEPRVIVISHAPRSLSRWMAALSGAALGTHVQGPRLRG- 347
441 IAL-----PMSADLKUYDVSHS---SMRAKN---GVAAGTGWMIYAPLITGL 483
348 EAQORVVPAGRNCTLO--GLAPGTAYLVTVTAAPRSGRESALSAACT 394
484 MADEKEIKIGEAStELDLGLPNTYTVYVYAMFGEASDPLTGQETT 532

RESULT 5

A45974
collagen alpha 1(XIV) chain precursor, short form 2 - chicken
N/Alternate names: undulin
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A45974; S30085; S22916; S17035; S20833
R.Giercke, D.R.; Foley, J.W.; Castagnola, P.; Gennart, M.; Dublet, B.; Cancedda, R.; Lin J. Biol. Chem. 268, 12177-12184, 1993
A/Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region
ns

A/Reference number: A45974; MUID:93280195; PMID:8505337
A/Accession: A45974
A/Status: preliminary
A/Molecule type: mRNA; protein
A/Residues: 1-1747 <GER>
A/Cross-references: UNIPROT:P32018; UNIPARC:UPI0000173C47
A/Experimental source: embryo skin
A/Note: sequence inconsistent with the nucleotide translation
A/Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBI:P.133365)
R.Apte, S.S.
submitted to the EMBL Data Library, March 1992
A/Reference number: S30085
A/Accession: S30085
A/Molecule type: mRNA
A/Residues: 1472-1660 <APR>
A/Cross-references: UNIPARC:UPI0000006A2; EMBL:X65122; NID:G62871; PID:CAA46238.1; PID:R.Trueb, J.; Trueb, B.
Eur. J. Biochem. 207, 549-557, 1992
A/Title: Type XIV collagen is a variant of undulin.
A/Reference number: S22916; MUID:9239443; PMID:1339349
A/Accession: S22916
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 286-494; 'Q', 496-834, 'A', 836-1119, 'K', 1122-1402, 1409-1439 <TRU>
A/Cross-references: UNIPARC:UPI0000173C48
R.Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemeayer, T.F.; van der Rest, M.; Mayne, Eur. J. Biochem. 201, 333-338, 1991
A/Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens
A/Reference number: S17035; MUID:92037585; PMID:1935930
A/Accession: S17035

A/Molecule type: protein
A/Residues: 1551-1570;1593-1639-1667 <GOR2>
A/Cross-references: UNIPARC:UPI0000173C49
A/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F.40-204/Domain: von Willebrand factor type A repeat homology <VWA1>
F.326-317/Domain: fibronectin type III repeat homology <FN3A>
F.326-409/Domain: fibronectin type III repeat homology <FN3A>
F.418-498/Domain: fibronectin type III repeat homology <FN3C>
F.507-591/Domain: fibronectin type III repeat homology <FN3D>
F.625-707/Domain: fibronectin type III repeat homology <FN3D>
F.716-798/Domain: fibronectin type III repeat homology <FN3D>
F.806-893/Domain: fibronectin type III repeat homology <FN3D>
F.924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>
F.1111-1152/Domain: non-collagenous NC4 #status predicted <NC4>
F.1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>
F.1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 17.2%; Score 369.5; DB 2; Length 1747;
Best Local Similarity 28.1%; Pred. No. 5.1e-18;
Matches 115; Conservative 64; Mismatches 159; Indels 71; Gaps 13;

27 PASAPRGDMLPDLSSASVSHYFRRFVGQVLPPLGTGALRASLVHGSRRPYTF 86
38 PAIA---DIVILVDGWSIGRNFRLVRLFLNLVSAFVSGSKTRVGLAQSGDPRIM 94
87 PFGQSSGEAODAVRASQRMGDHTGTALVYAKQFLPAEASGARPGVPKYLWVTDG 146

Db 95 HLNAYGKAVLDADNRNLPKGGNTLTGLATYILENCFPEEAGAPRGVSKITLITDCK 154

QY 147 SSDVPGPMPQELKDLGVTVFVSTGRGNFLESAASAPAEKHLHFV-DVDDLIHIVQ-- 203
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
C/Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C/Accession: I51027

Db 155 SQDDVTPPAKNLKEAGLELFAIGVKNADINELKEIASEPSTHYVNAVDFNMSIVEGL 214
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C/Accession: A40970

QY 204 -----ELRGSIIDAM-RPQQLHATEITSSGFPLAMPPLITADSG----- 241
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
C/Species: M.; Herpest, H.; Hummel, M.; Duertkop, H.; Tripflier, D.; Stein, H.; Schuppan, D.
J. Biol. Chem. 266, 17326-17332, 1991

Db 215 TRTVCSRVREOEKEIKGTITIASIGAPDLYTSDITARGFVSW-----THSPGKYEKRVV 270
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Reference number: A40970; MUID:9137351; PMID:1716629

QY 242 YYVLELVPSAQPGAPARQQLPGNATIMVAGLDPPDDYDALV-----PESN 288
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Accession: A40970

Db 271 YF-----PTRGQPEEVVVDSSSTAVLKNLMSLTIEYQIAVPAISYNAASEGLRGETT 324
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Molecule type: mRNA

QY 289 VLLLRPOLLRVTRPEEAGPERIVISHARPRLRVSNAPALSGAALGVHVGCPRLRG- 347
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Residues: 1-843 <JUS>

Db 325 LAL-----PMASDLKLVDSH---SWRAKMN---GVAAGATGVMLLAPLLEGL 367
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Cross-references: UNIPROT:Q05707; UNIPARC:UPI00000716A9; GB:M64108; NID:g3440081; PIDN
C/Keyword: glycoprotein

QY 348 EAQREVPAGRNCTTQ--GLAPGTAVLVTVTAAPNSGRESALSAACT 394
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
F/165-246/Domain: fibronectin type III repeat homology <FN3A>
F/355-338/Domain: fibronectin type III repeat homology <FN3B>
F/347-427/Domain: fibronectin type III repeat homology <FN3C>
F/436-520/Domain: fibronectin type III repeat homology <FN3D>
F/547-632/Domain: fibronectin type III repeat homology <FN3E>
F/541-723/Domain: fibronectin type III repeat homology <FN3F>
F/731-818/Domain: fibronectin type III repeat homology <FN3G>

Db 368 AADEKIKIGEASTLELDGLTNPTEYTVVYAMFGEASDPLTGOETT 416
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Accession: I51027

RESULT 6

I51027
type XII collagen alpha-1 chain - eastern newt (fragment)
C/Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C/Accession: I51027

R/Wei, Y.; Yang, E.V.; Klatt, K.P.; Taseava, R.A.
Dev. Biol. 168, 503-513, 1995

A/Title: Monoclonal antibody M12 identifies the urodele alpha 1 chain of type XII collag

A/Reference number: I51027; MUID:95246925; PMID:7729585

A/Accession: I51027

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-929 <MEI>

A/Cross-references: UNIPROT:Q91145; UNIPARC:UPI0000126D2E; EMBL:U19494; NID:g632647; PII
F/155-236/Domain: fibronectin type III repeat homology <3FR>
F/531-795/Domain: von Willebrand factor type A repeat homology <VMA3>

Query Match 12.9%; Score 278; DB 2; Length 929;

Best Local Similarity 24.0%; Pred. No. 7.2e-12;

Matches 87; Conservative 51; Mismatches 129; Indels 96; Gaps 7;

QY 22 AERGP-PASA-----PRGDLMLDSSASVSHYFSPRVGEFVGQVAPLPGTGLRAS 74
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 614 APDSPIPSGLDCTTKAQADIVLVDSWSIGRPNFYVNFISRVVEVFDISDRVQIA 673
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 75 LVHVGSRPYTEFPFGQHSGEAAQDAVRASAOQMDHTGTALVYAKEQLFAEASGARPG 134
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 674 VSGYSDPRTFEMGLNTHKTGKSLMDAVANLPYKGGVNTNSALKFLLENNFRGVGMREK 733
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 135 VPKVLVAVVTDGSSDPVGPMPQELKDLGVTVFVSTGRGNFLESAASAPAEKHLHFV- 193
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 734 ARKIALILTGTGKSQDDIVAPSKRYADEGIELYAVGIKNADENELKEIASDPDELVMYNA 793
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 194 DVDDLIHIVQELGSLDAMRPOQLHATEITSSGFPLAMPPLITADSGYVLELVPSAQ 253
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 794 DFEGLTNVAVDLLENVCNSVKG----- 815
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 254 GAARQQLPGNATIMVAGLDPPDDYDALVPESNVRLRPQLLRVTRPEEAGPERIVT 313
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 816 -----PG-----GLNP-----PSNLVT 827
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 314 SHAPRSLRVSNAPALSGAALGVHVGCPRLRGSEARVEVPAGRNCTTQGLAPGTAVL 373
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 828 SEPTPRSFRTVWPP--SQSVERFKVEYVYVAGGRQGEVYVVGRTQTTVLVGLKPEPEYV 885
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 374 VTV 376
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 886 VNV 888

RESULT 7

A40970
undulin 1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C/Accession: A40970

QY 204 -----ELRGSIIDAM-RPQQLHATEITSSGFPLAMPPLITADSG----- 241
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
C/Species: M.; Herpest, H.; Hummel, M.; Duertkop, H.; Tripflier, D.; Stein, H.; Schuppan, D.
J. Biol. Chem. 266, 17326-17332, 1991

Db 215 TRTVCSRVREOEKEIKGTITIASIGAPDLYTSDITARGFVSW-----THSPGKYEKRVV 270
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Reference number: A40970; MUID:9137351; PMID:1716629

QY 242 YYVLELVPSAQPGAPARQQLPGNATIMVAGLDPPDDYDALV-----PESN 288
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Accession: A40970

Db 271 YF-----PTRGQPEEVVVDSSSTAVLKNLMSLTIEYQIAVPAISYNAASEGLRGETT 324
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Molecule type: mRNA

QY 289 VLLLRPOLLRVTRPEEAGPERIVISHARPRLRVSNAPALSGAALGVHVGCPRLRG- 347
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Residues: 1-843 <JUS>

Db 325 LAL-----PMASDLKLVDSH---SWRAKMN---GVAAGATGVMLLAPLLEGL 367
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Cross-references: UNIPROT:Q05707; UNIPARC:UPI00000716A9; GB:M64108; NID:g3440081; PIDN
C/Keyword: glycoprotein

QY 348 EAQREVPAGRNCTTQ--GLAPGTAVLVTVTAAPNSGRESALSAACT 394
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
F/165-246/Domain: fibronectin type III repeat homology <FN3A>
F/355-338/Domain: fibronectin type III repeat homology <FN3B>
F/347-427/Domain: fibronectin type III repeat homology <FN3C>
F/436-520/Domain: fibronectin type III repeat homology <FN3D>
F/547-632/Domain: fibronectin type III repeat homology <FN3E>
F/541-723/Domain: fibronectin type III repeat homology <FN3F>
F/731-818/Domain: fibronectin type III repeat homology <FN3G>

Db 368 AADEKIKIGEASTLELDGLTNPTEYTVVYAMFGEASDPLTGOETT 416
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
A/Accession: I51027

Query Match 12.5%; Score 268.5; DB 2; Length 843;

Best Local Similarity 25.7%; Pred. No. 3e-11;

Matches 95; Conservative 53; Mismatches 152; Indels 69; Gaps 12;

QY 66 LGTGLRASLVHVGSRPYTEFPFGQHSGEAAQDAVRASAOQMDHTGTALVYAKEQLF 125
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 3 VQSEKTRIGLAQYSGDPRIEMHLNPFSTQDEVIEAVRNLPYKGGNTLTGLALNYIFENSF 62
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 126 AASAGAPGVPAVLVAVTVDGSSDPVGPMPQELKDLGVTVFVSTGRGNFLESAASAP 185
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 63 KEKAGSRVTSKIGITITDGTGSQDDIIPSRNLRREGVELFAIGVKNADINELKEIASDP 122
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 186 AEKHLHFV-DVDDLIHIVQELGSLDAMR-----PQQLHATEITSSGF 228
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 123 DETHVNVAVFELMHTVVESTLRITCSRVQEDREIKASAHATGPPTELTISEVARSF 182
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 229 RLAMPPLITADSG-----YYVLELVPSAQPGARQQLPGNATIMVAGLDPPDDY 280
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 183 MNW-----THAPGNVEKRVVY-----PTRGKRPDEVVVDGTVSSVLKNLMSLTIEYQ 232
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 281 VALVPESNVRLRPQLLRVTRPEE-AGPER-----IYSHARPRLRVSNAPALG 330
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 233 IAVF-----AIYHTASBGLRGTEITLALPMASDLIVDTENSMRVKMDAVFG 281
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 331 SAAALGVHVGCPRLRGSEA-QREVPAGRNCT--TLQGLAPGTAVLVTVTAFR----- 381
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 282 ---ASGYLLVAPLREGLAGDEKEMKIGETHRIDIELSGLLPNTBYTVTVYAMFGEASDP 338
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 382 -SGRESALS 389
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 339 VTGOETTLA 347
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
QY 347

RESULT 8

A33809
cartilage matrix protein precursor - chicken
C/Species: Gallus gallus (chicken)
C/Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 09-Jul-2004
C/Accession: A33809; A26364

R/Kise, I.; Deak, F.; Holloway Jr., R.G.; Dellus, H.; Mebuer, K.A.; Frimberger, E.; Arg
U. Biol. Chem. 264, 8126-8134, 1989

A/Title: Structure of the gene for cartilage matrix protein, a modular protein of the e.
grins, von Willebrand factor, complement factors B and C2, and epidermal growth factor.

A/Reference number: A33809; MUID:89255246; PMID:2542265

A/Accession: A33809

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-493 <KIS>

A/Cross-references: UNIPROT:P05099; UNIPARC:UPI000004FLBF; GB:X12346; GB:X12347; GB:X12

R:Argaves, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goetlinck, P.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 464-468, 1987
A>Title: Structural features of cartilage matrix protein deduced from cDNA.
A:Reference number: A26364; MUID:87092429; PMID:3025875
A:Accession: A26364
A:Molecule type: mRNA
A:Residues: 78-493 <ARG>
A:Cross-references: UNIPARC:UPI000004F1EE; GB:M14792; NID:g211545; PIDN:AAA48695.1; PID:
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repeat
F:37-204/Domain: von Willebrand factor type A repeat homology <VMA1>
F:225-260/Domain: EGF homology <EGF>
F:270-434/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 11.5%; Score 248.5; DB 2; Length 493;
Best Local Similarity 37.1%; Pred. No. 4e-10;
Matches 66; Conservative 22; Mismatches 83; Indels 7; Gaps 2;

25 GPAPASPRG-----DLMPFLDSSASVSHYEFGRVREFVQGVAPLPGLTGALRASLVHYV 78
257 GRTCSAGSGSGSALDVLFLIDGSKVRPBNFELVKFTNTQIVESLEVBKQAQVGLVQY 316
79 GSRPYTEPFPGQHSGEAAQDAVRASQRMGDHTHTGLALVYAKEQLFAEASGARPGVPKY 138
317 SSSVRGEPLGGQPKKKKDIKAAVKQMAVMEKGMTGQALKYLVDSFSIANGARPGVPKY 376
QY 139 LVMWTDGSSDPVPGPMQELKDIGTVFVYVSTGRGNFLELSAASAPAEKHLHFVDVD 196
DB 377 GIVFTDGRSDYITDAKAKKDGFMFVAVGVNAVEDELREIASEPVAEH-VFYRAD 433

RESULT 9
S66522
cartilage matrix protein precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S66522
R:Aszodi, A.; Hauser, N.; Studer, D.; Paulsson, M.; Hirtl, L.; Bosze, Z.
Eur. J. Biochem. 236, 970-977, 1996
A>Title: Cloning, sequencing and expression analysis of mouse cartilage matrix protein c
A:Reference number: S66522; MUID:96270751; PMID:8665920
A:Accession: S66522
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-500 <MS2>
A:Cross-references: UNIPROT:P51942; UNIPARC:UPI0000029480; EMBL:U55035; NID:g1163178; PI
C:Genetics:
A:Gene: CMP
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-500/Product: cartilage matrix protein #status predicted <MAT>
F:43-210/Domain: von Willebrand factor type A repeat homology <VMA1>
F:231-266/Domain: EGF homology <EGF>
F:277-441/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 11.3%; Score 243.5; DB 2; Length 500;
Best Local Similarity 34.3%; Pred. No. 9.3e-10;
Matches 60; Conservative 29; Mismatches 75; Indels 11; Gaps 2;

31 PRG-----DLMFLDSSASVSHYEFGRVREFVQGVAPLPGLTGALRASLVHYGSRP 82
34 PRGHLKTRPTDLVFDVDSRSRVPREFEKVFLQVIESLDVGNATRVGLVNTASTV 93
QY 83 YTEFPQHSSEGAADAVRASAQRMGDHTHTGLALVYAKEQLFAEASGAR--PGVPKYV 139
DB 94 KEFPPLRAHGSKSLQAVRRIQPLSTGWTGLALQFAIRKALSDAEGGARSPDISKTV 153
QY 140 VWVTDDGSSDPVPGPMQELKDIGTVFVYVSTGRGNFLELSAASAPAEKHLHFVD 194
DB 154 IVVTDGRPDSDVRSERASGIELFALGLGRVDRATLRQLASPEQDEHDVVE 208

RESULT 10
A37797

collagen alpha 3(VI) chain precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Aug-2004
C:Accession: A37797; A34270; A32674
R:Doliana, R.; Bonaldo, P.; Colombatti, A.
J. Cell Biol. 111, 2197-2205, 1990
A>Title: Multiple forms of chicken alpha3(VI) collagen chain generated by alternative sp
A:Reference number: A37797; MUID:91035630; PMID:1977751
A:Molecule type: mRNA
A:Residues: 1-253;312-321;434-453;504-518;635-655;704-717;832-853 <DOL>
A:Cross-references: UNIPROT:P15989; UNIPARC:UPI0000173C31; UNIPARC:UPI0000173C32; UNIPAR
M24282
R:Bonaldo, P.; Russo, V.; Buccicotti, F.; Doliana, R.; Colombatti, A.
Biochemistry 29, 1245-1254, 1990
A>Title: Structural and functional features of the alpha3 chain indicate a bridging role
A:Reference number: A34270; MUID:90212613; PMID:2322559
A:Accession: A34270
A:Molecule type: mRNA
A:Residues: 224-2871 <BON>
A:Cross-references: UNIPARC:UPI0000173C38; GB:M24282
A>Note: the authors translated the codon TTC for residue 1916 as Leu and TTC for residue
J. Biol. Chem. 264, 20235-20239, 1989
R:Bonaldo, P.; Colombatti, A.
A>Title: The carboxyl termini of the chicken alpha3 chain of collagen VI is a unique mo
A:Reference number: A32674; MUID:90062147; PMID:2584214
A:Accession: A32674
A:Molecule type: mRNA
A:Residues: 2151-2199;2792-3137 <BO2>
A:Cross-references: UNIPARC:UPI0000173C39; UNIPARC:UPI0000173C3A; GB:M24282
C:Genetics:
A:Introns: 30/1, 236/1; 437/1; 638/1; 838/1
C:Superfamily: collagen VI
C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; extracel
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-3137/Product: collagen alpha 3(VI) chain #status predicted <MAT>
F:26-204/Domain: non-collagenous #status predicted <NMC>
F:36-202/Domain: von Willebrand factor type A repeat homology <VM01>
F:39-404/Domain: von Willebrand factor type A repeat homology <VM02>
F:442-607/Domain: von Willebrand factor type A repeat homology <VM03>
F:642-807/Product: collagen type A repeat homology <VM04>
F:840-1004/Domain: von Willebrand factor type A repeat homology <VM05>
F:1033-1197/Domain: von Willebrand factor type A repeat homology <VM06>
F:1233-1400/Domain: von Willebrand factor type A repeat homology <VM07>
F:1439-1604/Domain: von Willebrand factor type A repeat homology <VM08>
F:1639-1804/Domain: von Willebrand factor type A repeat homology <VM09>
F:1839-2010/Domain: von Willebrand factor type A repeat homology <VM10>
F:2043-2378/Domain: collagenous #status predicted <COL>
F:2045-2047/Region: cell attachment (R-G-D) motif
F:2153-2155/Region: cell attachment (R-G-D) motif
F:2155-2161/Region: cell attachment (R-G-D) motif
F:2379-3137/Domain: non-collagenous #status predicted <CNC>
F:2405-2577/Domain: von Willebrand factor type A repeat homology <VM11>
F:2623-2806/Domain: von Willebrand factor type A repeat homology <VM12>
F:2803-2846/Domain: platelet glycoprotein Ib-like #status predicted <GPI>
F:2954-3039/Domain: fibronectin type III repeat homology <FN3>
F:3072-3122/Domain: animal knitin-type proteinase inhibitor homology <BPI>
F:201,2084,2436,2563,2683,2867,2920,3003/Binding site: carbohydrate (Asn) (covalent

Query Match 11.1%; Score 239; DB 2; Length 3137;
Best Local Similarity 33.5%; Pred. No. 2e-08;
Matches 76; Conservative 31; Mismatches 106; Indels 14; Gaps 6;

2 LPWTA-LGIALSLRLALRSAGRPPASAPR---GDLMFLDSSASVSHYEFGRVREF 56
DB 7 LPLAAILGLILS---GFSVGAGQ---QAAYNVAVADIIFLVDSWSIGKEHFLVREF 60
QY 57 VQGVAPLPGLTGALRASLVHYGSRPYTEFPQHSSEGAADAVRASAQRMGDHTHTGLA 116
DB 61 LVDVYKALDVGNDRFRALVQPSGNPHTEFQNTPTPSNODVLSHTANMPYMGSGSKTGG 120
QY 117 LVYAKQQLFAEASGAR--PGVPKYVVMWTDGSSDPVPGPMQELKDIGTVFVYVSTGRGN 174

Db 121 LEVLINHLTKAAGSRASEGVPIITVLTGQSDVDVAFSSVLKASHVMIAVGVODAV 180
QY 175 FLELSAASAPAEKHL-HFVVDVDELHIIVQELSGSLIDAMPQOLMA 220
181 EGLKEIASRPFDTLHFNLENFTALHGVIGDVLAVSTMTPTQAGA 227
RESULT 11
A37979
Cartilage matrix protein precursor - human
C:Species: Homo sapiens (man)
C:Date: 12-Jul-1991 #sequence, revision 12-Jul-1991 #text_change 09-Jul-2004
C:Accession: A37979; B37979
R:Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; Sh
J. Biol. Chem. 265, 19624-19631, 1990
A:Title: Structure and chromosomal location of the human gene encoding cartilage matrix
A:Reference number: A37979; PMID:9106056; PMID:2246248
A:Accession: A37979
A:Molecule type: DNA
A:Residues: 1-496 <JEN>
A:Cross-references: UNIPROT:P21941; UNIPARC:UPI000004F1ED; GB:J05667
A:Accession: B37979
A:Molecule type: mRNA
A:Residues: 157-290, 'L', 292-496 <JED>
A:Cross-references: UNIPARC:UPI000016A6E8; GB:M55683; GB:J05663; GB:J05667; NID:9180651;
C:Gene: GDB:CRTM
A:Map position: 1p35-1p35
A:Insertions: 32/1; 147/3; 222/1; 264/1; 403/1; 454/1; 481/1
C:Complex: homotrimer
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A rep
C:Keywords: glycoprotein; homotrimer
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-496/Product: cartilage matrix protein #status predicted <MAT>
F:29-206/Domain: von Willebrand factor type A repeat homology <WMA1>
F:227-262/Domain: EGF homology <EGF>
F:273-437/Domain: von Willebrand factor type A repeat homology <WMA2>
F:76,344/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:221-238,234-247,249-262/Dismulfide bonds: #status predicted
Query Match 10.8%; Score 233.5; DB 2; Length 496;
Best Local Similarity 35.0%; Pred. No. 4,7e-09;
Matches 62; Conservative 27; Mismatches 87; Indels 1; Gaps 1;
QY 34 DLMEFLIDSSASVSHYFERSRREFPGOLVAPLPLGTGALRLASLVHVGSRPTTPEPQGHSS 93
275 DLVFLIDGSKSVRPENELVKFTISQIVDTLDVSDKLAQVGVLYSSSVRQEFPLGFRHT 334
Db 94 GEAADAVRASAOQRMGDTHTGLALVYAKQLFAEASGARPGVPVLVWVTDDGSSDPVGP 153
335 KKDITKAIVRMMSYMEKQTMTCGALKTYLIDNSFTVSSARAGAKVGVLPFDGSSQDVI 394
QY 154 PMQELDLGVTVFIVSTGRGNFELSAAASAP-AEKHLHFVVDVDELHIIVQELSGSL 209
Db 395 AAKKAKDLGKMFVAVGVGNVDELREIASEPVAEHFYADFTINQIGKLOKTKI 451
RESULT 12
GHU3A
collagen alpha 3(VI) chain precursor [validated] - human
N:Contents: collagen alpha 3(VI) chain, splice form AS/N10(-)
C:Species: Homo sapiens (man)
C:Date: 21-Nov-1993 #sequence, revision 12-Nov-1999 #text_change 16-Aug-2004
C:Accession: A59140; S13679; S24465; A57083; S28776; S00245; C31952; C29848; S26510; S48
R:Chu, M.L.
submitted to Genbank, May 1998
A:Reference number: A59140
A:Accession: A59140
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-3176 <CHU>
A:Cross-references: UNIPROT:P12111; UNIPARC:UPI0000126D4F; GB:X52022; NID:93127925; PIDN

R:Chu, M.L.; Zhang, R.Z.; Pan, T.; Stokes, D.; Conway, D.; Kuo, H.J.; Glanville, R.; Ma
EMBO J. 9, 385-393, 1990
A:Title: Molecular structure of globular domains in the human type VI collagen alpha-3 cha
itors.
A:Reference number: S13679; PMID:90151612; PMID:1689238
A:Accession: S13679
A:Molecule type: protein
A:Residues: 1-30,237-313, 'CW', 318-322, 'AR', 326-1815, 'FD', 1818-1819, 'ID', 1822-3176 <CHS
A:Cross-references: UNIPARC:UPI0000173C14; EMBL:X52022; NID:93127925
A:Accession: S24465
A:Molecule type: mRNA
A:Residues: 1-30,237-313, 'CW', 318-322, 'AR', 326-1815, 'FD', 1818-1819, 'ID', 1822-3176 <CHS
A:Cross-references: UNIPARC:UPI0000173C28
A:Accession: S28776
A:Molecule type: mRNA
A:Residues: 32-126, 'AK', 129-136, 'L', 138-236 <ZA2>
A:Cross-references: UNIPARC:UPI0000173C28; GB:S49432; NID:9260296; PIDN:AA824261.1; PID
R:Chu, M.L.; Mann, K.; Deutzmann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.
Eur. J. Biochem. 168, 309-317, 1987
A:Title: Characterization of three constituent chains of collagen type VI by peptide se
A:Reference number: S00126; PMID:88029444; PMID:3655927
A:Accession: S00245
A:Molecule type: protein
A:Residues: 2024-2046; 2092-2156, 'R', 2203-2208, 'X', 2210-2221, 'X', 2213-2227, 2228-2251; 231
A:Cross-references: UNIPARC:UPI000016A705; UNIPARC:UPI0000173C29; UNIPARC:UPI0000173C2A
ID:91335034
A:Note: the mRNA portion of the sequence corresponds to residues 2092-2157
R:Chu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R.
J. Biol. Chem. 263, 18601-18606, 1988
A:Title: Amino acid sequence of the triple-helical domain of human collagen type VI.
A:Reference number: A31952; PMID:8906644; PMID:3198591
A:Accession: C31952
A:Molecule type: mRNA
A:Residues: 2038-2373 <CH4>
A:Cross-references: UNIPARC:UPI0000173C2E; GB:J04211; GB:M20778
A:Note: parts of this sequence were determined by protein sequencing
R:Weil, D.; Mattei, M.G.; Fassege, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deu
Am. J. Hum. Genet. 42, 435-445, 1988
A:Title: Cloning and chromosomal localization of human genes encoding the three chains
A:Reference number: A29848; PMID:88161046; PMID:3348212
A:Accession: C29848
A:Molecule type: mRNA
A:Residues: 2092-2151 <WE1>
A:Cross-references: UNIPARC:UPI000016A720; GB:M27449; NID:9291919; PIDN:AAA52057.1; PID
A:Note: part of this sequence was determined by protein sequencing
R:Jander, R.; Rautenberg, J.; Glanville, R.W.
Eur. J. Biochem. 133, 39-46, 1983
A:Title: Further characterization of the three polypeptide chains of bovine and human s
A:Reference number: S26506; PMID:83209648; PMID:6852033
A:Accession: S26510
A:Molecule type: protein
A:Residues: 'SAGVAVGV', <JAN>
A:Cross-references: UNIPARC:UPI0000173C2F
A:Note: this sequence cannot be reliably placed and probably represents the results fro
R:Keyer, U.; Poeschl, E.; Nischt, R.; Speck, U.; Pan, T.C.; Chu, M.L.; Timpl, R.
Eur. J. Biochem. 225, 573-580, 1994
A:Title: Recombinant expression and properties of the Kunitz-type protease-inhibitor mo
A:Reference number: S48709; PMID:95045506; PMID:7525281
A:Accession: S48709
A:Molecule type: mRNA
A:Residues: 'WRAMIFLLCLAGALAA', 3102-3176 <MAY>
A:Cross-references: UNIPARC:UPI0000173C30

A>Note: engineered sequence to allow isolation of the Kunitz-type proteinase inhibitor h R.Arnoux, B.; Merigau, K.; Saluajan, P.; Norris, F.; Norris, K.; Bjorn, S.; Olsen, O.; submitted to the Brookhaven Protein Data Bank, August 1994

A:Reference number: A52812; PDB:KNT

A:Contents: annotation: X-ray crystallography, 1.6 angstroms, residues 3106-3160

A>Note: engineered sequence expressed in *Saccharomyces cerevisiae* strain mt-663

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C) lines are 5-hydroxylated and subsequently O-glycosylated.

C:Comment: The fibronectin type III repeat homology domain may be released during process

C:Gene: GDB:COL6A3

A:Cross-references: GDB:119066; OMIM:120250

A:Map position: 2q37.3-2q37.3

C:Complex: type VI collagen is a heterotrimer of one alpha 1(VI) chain (see PIR:CGH1A), associations among trimer amino- and carboxyl-terminal domains (with disulfide bonds).

C:Function:

A:Description: structural component of extracellular tissue microfibrils associated with C:Superfamily: collagen VI

C:Keywords: alternative splicing; blocked amino end; cell binding; coiled coil; extrace

F:1-25/Domain: signal sequence #status predicted <Sig>

F:26-3176/Product: collagen alpha 3(VI) chain #status predicted <MAT1>

F:26-3037/Domain: amino-terminal nonhelical #status predicted <ANH>

F:26-30,237-3176/Product: collagen alpha 3(VI) chain, splice form A9/N10(-) #status pred

F:37-203/Domain: von Willebrand factor type A repeat homology <VM01>

F:240-405/Domain: von Willebrand factor type A repeat homology <VM02>

F:443-608/Domain: von Willebrand factor type A repeat homology <VM03>

F:637-802/Domain: von Willebrand factor type A repeat homology <VM04>

F:835-999/Domain: von Willebrand factor type A repeat homology <VM05>

F:1027-1191/Domain: von Willebrand factor type A repeat homology <VM06>

F:1231-1399/Domain: von Willebrand factor type A repeat homology <VM07>

F:1433-1599/Domain: von Willebrand factor type A repeat homology <VM08>

F:1637-1802/Domain: von Willebrand factor type A repeat homology <VM09>

F:1836-2005/Domain: von Willebrand factor type A repeat homology <VM10>

F:2038-2373/Region: interrupted helical

F:2040-2042/Region: cell attachment (R-G-D) motif

F:2136-2138/Region: cell attachment (R-G-D) motif

F:2148-2150/Region: cell attachment (R-G-D) motif

F:2154-2156/Region: cell attachment (R-G-D) motif

F:2370-2372/Region: cell attachment (R-G-D) motif

F:2370-3176/Domain: carboxyl-terminal nonhelical #status predicted <CNH>

F:2400-2571/Domain: von Willebrand factor type A repeat homology <VM11>

F:2617-2800/Domain: von Willebrand factor type A repeat homology <VM12>

F:2865-2986/Region: alanine/lysine/proline/threonine/valine-rich repeats

F:2987-3072/Domain: fibronectin type III repeat homology <FN3>

F:3111-3161/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:26/Modified site: pyroglutamate carboxylic acid (Gln) (in mature form) #status predicted

F:108,116,202,251,2079,2331,2558,2671,2861,3036/Binding site: carboxylate (Asn) (coval

F:2100,2206,2239,2316,2319/Modified site: 4-hydroxyproline (Pro) #status experimental

F:2103,2209,2212,2332,2337/Binding site: carboxylate (Lys) (covalent) #status experimen

F:2103,2209,2212,2332,2337/Modified site: 5-hydroxylysine (Lys) #status experimental

F:3111-3161,3120-3144,3136-3157/Disulfide bonds: #status predicted

Query Match 10.5%; Score 225.5; DB 2; Length 3176;

Best Local Similarity 24.8%; Pred. No. 1.8e-07;

Matches 81; Conservative 52; Mismatches 143; Indels 51; Gaps 8;

QY 34 DLMFLDSSASVSHYEFVRFEVQGVLPPLGTLALPASLVHVGSRPYTEPFQGHSS 93

DB 39 DILFLVSSWTIGSEHFLQVRELYVYVSLAVGENDPHFALQFNGNHTTEFLTLTYRT 98

QY 94 GEAADAVPASQRMGDTHGLLVYAKQLFAEASGAR--PGVPKVLVWVTGGSSDPV 151

DB 99 KOEVLISHINMSYIGTNGTGKGLFYIMQSHLTKAASRAQGVDPQIVLVLTGHSKDL 158

QY 152 GPPMQLKDLGVTVFVTSGRGNFLLSAASAPAEKHL-HFVDVDDHLHYVDELRSGL 210

DB 159 ALPSALSKADVNVFAIGVEDADGALKELASEPLMMHFNENFTSLDHYGNLVSCH 218

QY 211 DAMRPOOLATETTSGRFLAMPPLTLTADSGVYLVLEVSQAQGAARQOLPENATDWIM 270

DB 219 SSVSPERADDTLKD-----ITAQDSADITFLID----- 248

QY 271 AGLDEPTDYDALVPESSVRLRLPQILKRTTPPEAGERRIYSHARPSL-----RWSW 325

DB 249 -SSNVTGVSFNVFAIIVDFLNL--EKPLTQOIRVG---VVOFSDEPRMTSLDPTYSK 302

QY 326 APALGSAALGYHVOFGPLRGGAQRV 352

DB 303 AQVLGAVRLGF-----AGSELANI 322

RESULT 13

T46488

hypothetical protein DKFZp434J065.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004

C:Accession: T46488

R:UniProt, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

A:Submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23035

A:Accession: T46488

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-741 <AAA>

A:Cross-references: UNIPROT:Q8NDB6; UNIPARC:UPI000016ACGB; EMBL:AL137638

A:Experimental source: adult testis; clone DKFZp434J065

C:Genetics:

A>Note: DKFZp434J065.1

Query Match 10.2%; Score 220; DB 2; Length 741;

Best Local Similarity 31.9%; Pred. No. 7.1e-08;

Matches 59; Conservative 36; Mismatches 84; Indels 6; Gaps 4;

QY 34 DLMFLDSSASVSHYEFVRFEVQGVLPPLGTLALPASLVHVGSRPYTEPFQGHSS 93

DB 440 DILFLVSSWTIGSEHFLQVRELYVYVSLAVGENDPHFALQFNGNHTTEFLTLTYRT 98

QY 94 GEAADAVPASQRMGDTHGLLVYAKQLFAEASGAR--GVPKVLVWVTGGSSDPV 149

DB 500 AKDMKAV-AHKYVKGSGMTLAKHMERSTFGEGARPLSTRPRAIVFTGRRAD 558

QY 150 PVGPPMQLKDLGVTVFVTSGRGNFLLSAASAPAEKHLFV-DVDDHLHYVDELRSGL 208

DB 559 DVEWASAKKANGITMVAVGKALBEDELASEPTKHLFLYADDFSTMDISRLKKG 618

QY 209 IUDAM 213

DB 619 ICEAL 623

RESULT 14

A55348

integrin alpha-1 - chicken (fragment)

C:Species: Gallus gallus (chicken)

C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004

C:Accession: A55348

R:UniProt, A.; Briesewitz, R.; Bank, I.; Marcantoni, E.B.

J. Biol. Chem. 269, 22811-22816, 1994

A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1

A:Reference number: A55348; MUID:94357930; PMID:7521332

A:Accession: A55348

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <KER>

A:Cross-references: UNIPROT:Q42094; UNIPARC:UPI000017A14E; GB:U10114

F:55-230/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 9.7%; Score 209; DB 2; Length 272;

Best Local Similarity 33.6%; Pred. No. 1.2e-07;

Matches 73; Conservative 31; Mismatches 85; Indels 28; Gaps 10;

QY 34 DLMFLDSSASVSHYEFVRFEVQGVLPPLGTLALPASLVHVGSRPYTEPFQGHSS 93

DB 57 DIVIYLDGNSI--YPMESVTAFNLNLMIDIGFQQTGVIGVIGQTVVHFEFLTYST 114

QY 94 GEAADAVASQRMG-DTHGTALVYAKEQLFAEASGARPGVKLVWVTDGSSDPVG 152
DB 115 TEEVMDALRIQRCGTMTALGIDTAHEEAFTENAGARGVQKVMVITVDGESH--N 172
QY 153 PMOEL-----KDLGTVFIV-----STGRGNFL-----ELSAASAPAEKHLHFVDVD-- 197
DB 173 YRLQEVIDKCEDENIQRFAIALIGSYSRNLSTKEFVEIKSIASKPTEK--HFFNVSD 230
QY 198 --LHIIVQELRGSII--LDAMPPOOLHA--TEITSSGF 228
DB 231 LALVTIVEALGERIFALEATTDQAASFEMWSQAGF 267

RESULT 15

A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Brileswitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; PMID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Cross-references: UNIPROT:P56199; UNIPARC:UPI000012D9EB
A:Experimental source: hepatoblastoma cell line HepG2
A>Note: sequence extracted from NCBI backbone (NCBI:P124326)
F:142-317/Domain: von Willebrand factor type A repeat homology <WMA1>

Query Match 9.7%; Score 208; DB 2; Length 1151;
Best Local Similarity 27.7%; Pred: 8.9e-07;
Matches 78; Conservative 44; Mismatches 104; Indels 56; Gaps 11;

QY 34 DLMFLDSSASVSHYEFGRVREFVQLVAPLPGTALRASLVHSGRPYEPFGQHS 93
DB 144 DIVIVLDGNSNI--YPMDSVTAFLNDLKRMDIGPKQTQVGIQYGENVTHEFNLYSS 201
QY 94 GEAADAVASQRMG-DTHGTALVYAKEQLFAEASGARPGVKLVWVTDGSSD--P 150
DB 202 TEEVMDALRIQRCGTMTALGIDTAHEEAFTENAGARGVQKVMVITVDGESHNR 261
QY 151 VCPMDELKDIGTVFIV-----STGRGNFL-----ELSAASAPAEKHLHFVDVD-- 197
DB 262 LKKVIODDCEDENIQRFAIALIGSYSRNLSTKEFVEIKSIASKPTEK--HFFNVSD 319
QY 198 LHIIVQELRGSII--LDAMPPOOLHA--TEITSSGFRIAMPPLLTADSGYVLELVSAOP 253
DB 320 LVTIVKTIGERIIFALEATTDQAASFEMWSQAGF-----SAHY----- 358
QY 254 GAARQQLPGNATDWIAGIDPPTDYDALVPESNRLARPQ 295
DB 359 -----SQDWMLGAVGAYDWNQVVMQKASQIITPR 389

Search completed: February 13, 2006, 07:49:18
Job time : 63.2107 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:33:59 ; Search time 329.326 Seconds
(without alignments)
893.868 Million cell updates/sec

Title: US-10-699-035a-6

Perfect score: 2154
Sequence: 1 MPMWTALGVALLSLRLALARS.....RRPRPVPRAPTPGTASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2130.5	98.9	445	2	Q6PCB0 HUMAN
2	1657.5	76.9	415	2	Q8R225 MOUSE
3	1653.5	76.8	415	2	Q923K3 MOUSE
4	1653.5	76.8	415	2	Q8C007 MOUSE
5	1643.5	76.3	415	2	Q642A6 RAT
6	1295.5	60.1	281	2	Q715D7 HUMAN
7	1058.5	49.1	233	2	Q9H6J5 HUMAN
8	782.5	36.3	203	2	Q8VDV9 MOUSE
9	684	31.8	505	2	Q4SCD1 TETNG
10	437.5	20.3	3124	1	COCA1 CHICK
11	424	19.7	2884	2	Q5VYK2 HUMAN
12	424	19.7	3063	1	Q5VYK1 HUMAN
13	421.5	19.6	3063	1	COCA1 HUMAN
14	411.5	19.1	1723	1	Q4SD22 TETNG
15	407	18.9	3119	1	COCA1 MOUSE
16	401	18.6	517	2	Q43853 HUMAN
17	392.5	18.2	1259	2	Q4RP12 TETNG
18	392.5	18.2	1284	2	Q4VXQ5 HUMAN
19	392.5	18.2	1284	2	Q4VXQ5 HUMAN
20	392.5	18.2	1297	2	Q4VXQ4 HUMAN
21	392.5	18.2	1329	1	K1510 HUMAN
22	392	18.2	637	2	Q81VX1 HUMAN
23	387	18.0	1117	2	Q4RXN8 TETNG
24	383	17.8	2225	2	Q4SXK3 TETNG
25	382.5	17.7	2944	1	CO7A1 CHICK
26	380.5	17.7	1888	1	COEAI HUMAN
27	380	17.6	1253	2	Q97566 CANFA
28	373	17.3	2936	2	Q7YRK8 CANFA
29	360	16.7	2944	2	Q63870 MOUSE
30	357	16.6	1472	2	Q902A0 CHICK
31	339	15.7	1797	1	COEAI_MOUSE

32	327.5	15.2	1557	2	Q4SH63 TETNG	Q4SH63 tetraodon n
33	324	15.0	1796	1	COEAI_HUMAN	Q05707 homo sapien
34	322.5	15.0	534	2	Q4G0W3 HUMAN	Q4G0W3 homo sapien
35	320.5	14.9	1816	2	Q6NT15 BRARE	Q6NT15 brachydanto
36	292.5	13.6	1060	2	Q4SG03 TETNG	Q4SG03 tetraodon n
37	283.5	13.2	639	1	COCA1_RABIT	Q28902 oryctolagus
38	278	12.9	929	1	COCA1_MOTIT	Q91145 notophthalm
39	266	12.3	624	1	MATN4_MOUSE	Q89029 mus musculu
40	259	12.0	622	1	MATN4_HUMAN	Q95460 homo sapien
41	256	11.9	839	2	Q6P3N7 XENTR	Q6P3N7 xenopus tro
42	254	11.8	647	2	Q4S2X7 TETNG	Q4S2X7 tetraodon n
43	253.5	11.8	451	2	Q8N2G3_HUMAN	Q8N2G3 homo sapien
44	253.5	11.8	956	1	MATN2_HUMAN	Q00339 homo sapien
45	253	11.7	688	2	Q4T0K3 TETNG	Q4T0K3 tetraodon n

ALIGNMENTS

RESULT 1
Q6PCB0 HUMAN
ID Q6PCB0 HUMAN PRELIMINARY; PRT; 445 AA.
AC Q6PCB0
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein, isoform 1 (WARP).
GN Name=WARP; ORFNames=RP4-758J18.11-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Krausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Krausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywicki M.T., Skalska U., Smalins D.E.,
RA Scherf A., Schein J.E., Jones S.J.W., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX Strausberg R.;
RA Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Hall R.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC059409; AAHS9409.1; -, mRNA.
DR EMBL: AL391244; CA122657.1; -, Genomic DNA.
DR Ensembl: ENSG00000179403; Homo sapiens.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR002035; WWF_A.
DR Pfam: PF00041; Fn3; 2.
DR Pfam: PF00092; VWA; 1.

DR	PRINTS: PR00453; VMFADOMAIN.
DR	SMART; SMO0060; FN3; 2.
DR	SMART; SMO0327; VMA; 1.
DR	PROSITE; PS50853; FN3; 2.
DR	PROSITE; PSS0234; VMFA; 1.
SQ	SEQUENCE 445 AA; 46804 MW; D9FEDE9C0A4DEBD F CRC64;
Query Match	98.9%; Score 2130.5; DB 2; Length 445;
Best Local Similarity	93.9%; Pred. No. 1,8e-137;
Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1	
OY	1 MLPTWALGLALSRLALARGSGERGPPAPAGDMLFLDSSASVHSFRERVGQL 60
Dd	1 MLPWTALGLALSRLALARSGARGPAPAGDMLFLDSSASVHSFRERVGOL 60
OY	61 VAPPLIGTGLTALSLVHVSRRPYTERPFQSHSGEAAODAVASAQRMDTHTGALYYA 120
Dd	61 VAPPLIGTGLTALSLVHVSRPYTERPFQSHSGEAAODAVASARMDHTHGALYYA 120
OY	121 KEQLFAEASGARGPVKVLVMYTGGSDVPDPMEQLKDLYGVTFIVSTGTGNFIETSA 180
Dd	121 KEQLFAEASGARGPVKVLVMYTDDGSSDPVPMPMELDKLGYTVFIVSTGTGNFIETSA 180
OY	181 AASAPAEKLHHVDVDDHLIYOELRGSLLDMRPQQLATETITSSGFLLAMPPLLTADS 240
Dd	181 AASAPAEKLHHVDVDDLHIYOELRGSLLDMPRPOQLATEITSSGFLLAMPPLLTADS 240
OY	241 GYVVLELVESAOPGAARRQQLNGNATDWIMAGLDPTDVAVALVESNRLLRPOLLRYR 300
Dd	241 GYVVELVESAPPGAARRQQLNGNATDWIMAGLDPTDVAVALVESNRLLRPOLLRR 300
OY	301 TR-----PEEGAPERIVISHARPSSLRVSNAPALGSAA 333
Dd	301 TRPGAGCPGASGESGACGAPTQLAAPAEAGEPERIYISHARPSSLRVSNAPALGSAA 360
OY	334 ALGYVVOGPELRGSEAGORVEVPAGRNCITLQGLAPCTALVMVTAAFRSGRSALSAKAC 393
Dd	361 ALGYVVOGPELRGGEQREVPAGRNCTTLQGLAPCTALVTVTAAFRSGRSALSAAKAC 420
OY	394 TPDDGRPRPRPVPRAPTPTGTAAREP 418
Dd	421 TPDGPRPRPRPVPRAPTPTGTAREP 445
RESULT 2	
OBR225 MOUSE	PRT; 415 AA.
ID OBR225_MOUSE PRELIMINARY;	
AC OBR225_	
DT 01-JUN-2002 (TREMBLrel. 21, Created)	
DT 01-JUN-2002 (TREMBLrel. 21, Iast sequence update)	
DT 13-SEP-2005 (TREMBLrel. 31, Iast annotation update)	
DE Von Willebrand factor A domain-related protein (Mus musculus 11 days pregnant adult female ovary and uterus cDNA, RIKEN full-length enriched library, Clone:5031410123 product:VON WILLEBRAND FACTOR A-RELATED PROTEIN homolog) .	
DE Name=Wval1;	
GN Mus musculus (Mouse) .	
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;	
OC Murioidea; Muridae; Murinae; Mus.	
OX NCBI_TaxId=10090;	
RN [1]	
RP NUCLEOTIDE SEQUENCE.	
RC STRAIN=Czech II, and FVB/N;	
RC TISSUE=Colon, and	
RC Mammary tumor metastatized to lung. Tumor arose spontaneously;	
RX MEDLINE=2238857; PubMed=12477932; DOI=10.1073/pnae.242603899;	
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,	
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuetz G.D.,	
RA Altschul S.F., Zeeberg B., Burow K.H., Schefer C.F., Bhat N.K.,	
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,	
RA Dietchenko L., Martusina K., Farmer A.A., Rubin G.M., Hong L.,	
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.J., Scheetz T.E.,	

RA Brownstein M.J., Usdin T.B., Toshhyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hader S.E., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shechenko Y., Boulford G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywiński M.I., Skalska U., Smallue D.E.,
RA Schenck A., Schein J.E., Jones S.J.W., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN NCBI/GenBank database.
RN STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RG NIH MCC Project;
RL Submitted (Apr-2002) to the EMBL/Genbank/DBJ databases.
[3]
RN NCBI/GenBank database.
RN STRAIN=C57BL/6j; TISSUE=Ovary and uterus;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
[4]
RN NCBI/GenBank database.
RN STRAIN=C57BL/6j; TISSUE=Ovary and uterus;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Klehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schrödl L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
RA Blake J., Hoffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Guatlichin S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima Y., Mazzarelli V., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storck K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawasaki H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[5]
RN NCBI/GenBank database.
RN STRAIN=C57BL/6j; TISSUE=Ovary and uterus;
RC The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[6]
RN NCBI/GenBank database.
RN STRAIN=C57BL/6j; TISSUE=Ovary and uterus;
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Komoto H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[7]
RN NCBI/GenBank database.
RN STRAIN=C57BL/6j; TISSUE=Ovary and uterus;
RC MEDLINE=20530913; PubMed=11076661; DOI=10.1101/gr.152600;
RX Shibata K., Itoh M., Aizawa K., Nagaoaka S., Saeki N., Carninci P.,
RA

RA Kono H., Akiyama J., Nishi K., Kiteumai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Ozawa K., Tanaka T., Matsunuma S., Kawai J.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunuma S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-Format
sequencing pipeline with 384 multicapillary sequencer".
RL Genome Res. 10:1757-1771(2000).
RN [8]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Ohnato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Taya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [9]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Director MGC Project;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC02619; AAK2619.1; -; mRNA.
DR EMBL: AK077240; BAC36703.1; -; mRNA.
DR EMBL: BC036166; AAK36166.1; -; mRNA.
DR HSSP: P18614; 1MHP.
DR Ensembl: ENSMUSG00000042116; Mus musculus.
DR MGI: MGI:2179729; Vwal.
DR GO: GO:0005615; C:extracellular space; TAS.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR002035; VWF_A.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00092; VWA; 1.
DR PRINTS: PRO0453; VWFADOMAIN.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS50853; FN3; 2.
DR PROSITE: PS50234; VWF; 1.
DR PROSITE: PS50234; VWF; 1.
SQ SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;
Query Match 76.8%; Score 1657.5; DB 2; Length 415;
Best Local Similarity 77.8%; Pred. No. 3.8e-105;
Matches 325; Conservative 31; Mismatches 59; Indels 3; Gaps 1;
QY 1 MLFWTALGLALSLRLALARSGERPPASAPRGDLMLFLDSSASVSHYEFSSRVREFVQGL 60
DB 1 MLFWTALGLALSLRLALARSGERPPASAPRGDLMLFLDSSASVSHYEFSSRVREFVQGL 60
QY 61 VADPLGTGALRASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
DB 61 VADPLGTGALRASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
QY 61 VAIMSGPGLRLASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
DB 61 VAIMSGPGLRLASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
QY 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
DB 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
QY 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
DB 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
QY 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
DB 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
QY 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
DB 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
QY 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
DB 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
QY 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
DB 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
QY 301 TPREEAGPERIVISHAPPSRLRVSAPALGSAALGYHVGCPRLRGEAQRVEVPAGRNC 360
DB 301 TPREEAGPERIVISHAPPSRLRVSAPALGSAALGYHVGCPRLRGEAQRVEVPAGRNC 360

DB 301 TLQEEAGPERIVISHAPPSRLRVSAPALGSAALGYHVGCPRLRGEAQRVEVPAGRNC 360
QY 361 TTLQGIAPGTAVLVYTAAFRSRGESALSAKACTPDGPRPRPVPAPPTGASREP 418
DB 361 TTLQGIAPGTAVLVYTAAFRSRGESALSAKACTPDGPRPRPVPAPPTGASREP 418
QY 361 TTLQGIAPGTAVLVYTAAFRSRGESALSAKACTPDGPRPRPVPAPPTGASREP 418
DB 361 TTLQGIAPGTAVLVYTAAFRSRGESALSAKACTPDGPRPRPVPAPPTGASREP 418
RESULT 3
Q923K3 MOUSE PRELIMINARY; PRT; 415 AA.
ID Q923K3 MOUSE PRELIMINARY; PRT; 415 AA.
AC Q923K3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Von Willebrand factor A-related protein.
GN Name=Vwal; Synonyms=4932416A1IRik, Warp;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22057805; PubMed=12062410; DOI=10.1016/S0014-5793(02)02579-6;
RA Fitzgerald J., Ting S.T., Bateman J.F.;
RT "Warp a new member of the von Willebrand factor A-domain superfamily
of extracellular matrix proteins".
RL FEBS Lett. 517:61-66(2002).
DR EMBL: AY030094; AAK36350.1; -; mRNA.
DR HSSP: P18614; 1MHP.
DR Ensembl: ENSMUSG00000042116; Mus musculus.
DR MGI: MGI:2179729; 4932416A1IRik.
DR MGI: MGI:2179729; Vwal.
DR GO: GO:0005615; C:extracellular space; TAS.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR002035; VWF_A.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00092; VWA; 1.
DR PRINTS: PRO0453; VWFADOMAIN.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS50853; FN3; 2.
DR PROSITE: PS50234; VWF; 1.
DR PROSITE: PS50234; VWF; 1.
SQ SEQUENCE 415 AA; 44737 MW; C3344AECB3FDE431 CRC64;
Query Match 76.8%; Score 1653.5; DB 2; Length 415;
Best Local Similarity 77.5%; Pred. No. 7.1e-105;
Matches 324; Conservative 32; Mismatches 59; Indels 3; Gaps 1;
QY 1 MLFWTALGLALSLRLALARSGERPPASAPRGDLMLFLDSSASVSHYEFSSRVREFVQGL 60
DB 1 MLFWTALGLALSLRLALARSGERPPASAPRGDLMLFLDSSASVSHYEFSSRVREFVQGL 60
QY 61 VADPLGTGALRASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
DB 61 VADPLGTGALRASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
QY 61 VAIMSGPGLRLASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
DB 61 VAIMSGPGLRLASLVHVSQRPTEPFPGQSSGEAAQDAVRAAQMGTHTGLALVYA 120
QY 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
DB 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
QY 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
DB 121 KEOLFPAEAGARPGVPRKVLVWVTGSSDPVGPMEKLDLGVTFIVSTGRGNFLFLSA 180
QY 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
DB 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
QY 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
DB 181 AASAPAEKHLHFVVDVLIIVQELRGSILDMARPOOLHATEITSSGFRILAMPPLTLADS 240
QY 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
DB 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
QY 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
DB 241 GYVVLDELVSAGPAGARROOLPGENATDWIAGLDPTDIDVALVPSNVLLRPQLILVR 300
QY 301 TPREEAGPERIVISHAPPSRLRVSAPALGSAALGYHVGCPRLRGEAQRVEVPAGRNC 360
DB 301 TPREEAGPERIVISHAPPSRLRVSAPALGSAALGYHVGCPRLRGEAQRVEVPAGRNC 360

OY 361 TTLOGIAPGTAVLVVTTAAFRSGRESALSAXACTPDGPRPRPVPRAPTPGTASREP 418
DB 361 TTVQGLTCTTIVLVVTTAAFRSGRORALSAXACTASGARTR--APQSMRPEAGPREP 415

RESULT 4
O8C007_MOUSE PRELIMINARY: PRT: 415 AA.
AC O8C007;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone 4932416A11 product: VON WILLEBRAND FACTOR A-RELATED PROTEIN homolog.
GN Name=Vwaf1; Synonyms=4932416A11R1K;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
[1]
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99379253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085560; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I., Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochi H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schiml L.M., Stabl P., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gasterich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima U., Mazzarelli J., Mombarts P., Norioka P., Ring B., Ringwald M., Rodriguez I., Sakomura N., Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and sequencing of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RX

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Saeki N., Carninci P., Kono H., Akiyama Y., Nishi K., Katsunari T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi U., Aizawa K., Akimura T., Aikawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hirozaki T., Hirozaki T., Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Kondo M., Koyama S., Kurihara C., Matsuyama T., Miyazaki A., Muraoka M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T., Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (Jul-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK030019; BAC26739.1; -; mRNA.
DR HSSP; P18614; 1MHP.
DR Ensembl; ENSMUSG0000042116; Mus musculus.
DR MGI; MGI:2179729; 4932416A11R1K.
DR MGI; MGI:2179729; Vwaf1.
DR GO; GO:0005615; Cytoplasmic space; TAS.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS00234; VWF_A; 1.
SQ SEQUENCE 415 AA; 44681 MW; CE137963B76834FE CRC64;

Query Match 76.8%; Score 1653.5; DB 2; Length 415;
Best Local Similarity 77.5%; Pred. No. 7.1e-105;
Matches 324; Conservative 31; Mismatches 60; Indels 3; Gaps 1;

OY 1 MLPTALGATSLRLALRSGRGPASAPPGDMLFLDSSASVSHFEFSVRFEVQGL 60
DB 1 MLPTALGATSLRLALRSGRGPASAPPGDMLFLDSSASVSHFEFSVRFEVQGL 60
OY 61 VAPLPLGTGALRASLVHVGSRPYTEPFQGHSSGEAADAVALASQRMGDTHTGLALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTEPFQGHSSGEAADAVALASQRMGDTHTGLALVYA 120
OY 121 KEQLFAEASGARPGPKLVWTTDGGSSDPVGPQELKDLGVYFVYISTGAGNLEISA 180
DB 121 KEQLFAEASGARPGPKLVWTTDGGSSDPVGPQELKDLGVYFVYISTGAGNLEISA 180
OY 121 KEQLFAEASGARPGPKLVWTTDGGSSDPVGPQELKDLGVYFVYISTGAGNLEISA 180
DB 121 KEQLFAEASGARPGPKLVWTTDGGSSDPVGPQELKDLGVYFVYISTGAGNLEISA 180
OY 181 AASPAEKHILHVVDDDDHIIYQELRGSLDAMRQQLAHTEITSSGRIAPPLITADS 240
DB 181 AASPAEKHILHVVDDDDHIIYQELRGSLDAMRQQLAHTEITSSGRIAPPLITADS 240
OY 241 GYVLELVPSAOPGAARQQLPGNATDWIWAGLDPTDYVALVESNVRLPQILRYR 300
DB 241 GYVLELVPSAOPGAARQQLPGNATDWIWAGLDPTDYVALVESNVRLPQILRYR 300
OY 301 TRPEAGFRIVYISARPRSLRVSWAPALGSAALGVYQGGPLRGGEARVYVAGANC 360
DB 301 TRPEAGFRIVYISARPRSLRVSWAPALGSAALGVYQGGPLRGGEARVYVAGANC 360
OY 361 TTLOGIAPGTAVLVVTTAAFRSGRESALSAXACTPDGPRPRPVPRAPTPGTASREP 418
DB 361 TTVQGLTCTTIVLVVTTAAFRSGRORALSAXACTASGARTR--APQSMRPEAGPREP 415

RESULT 5
0642A6 RAT PRELIMINARY; PRT; 415 AA.
ID 0642A6 RAT PRELIMINARY; PRT; 415 AA.
AC 0642A6; 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DR Von Willebrand factor A domain-related protein.
GN Name=RGD1311476_predicted;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Carninci P., Prange C.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RX NIH MGC Project;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC081983; AAH81983.1; -, mRNA.
DR InterPro, IPR003961; FN_III.
DR InterPro, IPR002035; VWF_A.
DR Pfam, PF00041; FN3; 2.
DR Pfam, PF00092; VWA; 1.
DR PRINTS, PRO00453; VWFADOMAIN.
DR SMART, SM00060; FN3; 2.
DR SMART, SM00327; VWA; 1.
DR PROSITE, PS00853; FN3; 2.
DR PROSITE, PS0234; VWF; 1.
SQ SEQUENCE 415 AA; 44832 MW; B3A0A0268D806D4E CRC64;
Query Match 76.3%; Score 1643.5; DB 2; Length 415;
Best Local Similarity 77.8%; Pred. No. 3.5e-104;
Matches 325; Conservative 31; Mismatches 59; Indels 3; Gaps 1;

QY 181 AASAPAEKHLHFVDVDDLIITVOELRGSIIDAMRPOQLATEITSSGFLPAMPPLITADS 240
DB 181 AASAPAEKHLHFVDVDDLIITVOELRGSIIDAMRPOQLATEITSSGFLPAMPPLITADS 240
QY 241 GYYVLELVPSAOPGARRQQLPGNATDWMAGIDPDTVDVALVPSNRRLRPOLLRYR 300
DB 241 GYYVLELVPSGKLAATRRQQLPGNATSWMTDNDPTDYEVSLLPESNRRLRPOLLRYR 300
QY 301 TRPEEAGEPRIVISHAPRSRLSVSWAPALGAAAGYHVFQPLRGGEAQRVEVPAGNRC 360
DB 301 TQEEAGEPRIVISHTRPSRLSVSWAPALGPDSTIGYLVQGLPQGSLEHYEVPAGQNS 360
QY 361 TTLQGLAETAVLVVYTAARFSGRESALSAKACTDGPERRPRRPVPRATPTGTASEP 418
DB 361 TTLQGLTPTCTVLTVYTAARFSGRORALSAKCTASGERIR--VQAMRPAGALEP 415
RESULT 6
07LSD7 HUMAN PRELIMINARY; PRT; 281 AA.
ID 07LSD7 HUMAN PRELIMINARY; PRT; 281 AA.
AC 07LSD7;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DR WARP protein (Fragment).
GN Name=WARP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
OC Homo.
OX NCBI_Taxid=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Carninci P., Prange C.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX NIH MGC Project;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC003543; AAH03543.2; -, mRNA.
DR InterPro, IPR003961; FN_III.
DR Pfam, PF00041; FN3; 2.
DR SMART, SM00060; FN3; 1.
DR PROSITE, PS00853; FN3; 2.
FT NON_TER 1 1
SQ SEQUENCE 281 AA; 29628 MW; 350CC34590791BD4 CRC64;
Query Match 60.1%; Score 1295.5; DB 2; Length 281;
Best Local Similarity 90.4%; Pred. No. 1.4e-80;
Matches 254; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

QY 165 VPIVSTGRGNFLLSAASAPAEKHLHFVDVDDLIITVOELRGSIIDAMRPOQLATEIT 224

```
Db      1 VIVTSGRGNFLELSAASAPAKHHPVDVDDLIITVDEKSIIDANKPQQLHATEIT 60
Qy      225 SSGFRIAMPPLLTADSGYVLELVPSAOPGAARQQLPGNATDWMAGIDPTDYVALV 284
Db      61 SSGFRIAMPPLLTADSGYVLELVPSAOPGAARQQLPGNATDWMAGIDPTDYVALV 120
Qy      285 PSSNVLLRPQLIRVTR-----PEKGPRIIVISHAR 317
Db      121 PSSNVLLRPQLIRVTRPGEAGPGASGPGAGAPPTQALPABEEAGPERIVISHAR 180
Qy      318 PSLRVSMAIPALGSAALGYHVOFGPLRGEAORVAVPAGRNCTTQGLAPGAYLVVYT 377
Db      181 PSLRVSMAIPALGSAALGYHVOFGPLRGEAORVAVPAGRNCTTQGLAPGAYLVVYT 240
Qy      378 AAFRSGRESALSAACTPDGPRPRPVPRAPPTGTASREP 418
Db      241 AAFRSGRESALSAACTPDGPRPRPVPRAPPTGTASREP 281
```

```
RESULT 7
QSH6J5_HUMAN
ID QSH6J5 HUMAN PRELIMINARY; PRT; 233 AA.
AC QSH6J5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FLJ22215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Homo.
OC NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Natanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsushi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Iino T., Sugano S.,
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025868; BAB15264.1; -, mRNA.
DR HSSP; Q96K27; 1FNA.
DR Ensembl; ENSG00000179403; Homo sapiens.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS50853; FN3; 2.
SQ SEQUENCE 233 AA; 24473 MW; B2CC118FCSB7EC50 CRC64;
```

```
Query Match 49.1%; Score 1058.5; DB 2; Length 233;
Best Local Similarity 88.4%; Pred. No. 1.8e-64;
Matches 206; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

Qy      213 MRPQQLHATEITSSGFRIAMPPLLTADSGYVLELVPSAOPGAARQQLPGNATDWMAG 272
Db      1 MRPQQLHATEITSSGFRIAMPPLLTADSGYVLELVPSAOPGAARQQLPGNATDWMAG 60
Qy      273 LDDPTDYVALVPSNVLLRPQLIRVTR-----PEE 305
Db      61 LDDPTDYVALVPSNVLLRPQLIRVTRPGEAGPGASGPGAGAPPTQALPABEE 120
Qy      306 AAPERIVISHARPRSLRVSWAPALGSAALGYHVOFGPLRGEAORVAVPAGRNCTTQGL 365
Db      121 AAPERIVISHARPRSLRVSWAPALGSAALGYHVOFGPLRGEAORVAVPAGRNCTTQGL 180
Qy      366 LAPGTAIVTVTAAPRSGRESALSAACTPDGPRPRPVPRAPPTGTASREP 418
Db      181 LAPGTAIVTVTAAPRSGRESALSAACTPDGPRPRPVPRAPPTGTASREP 233

RESULT 8
Q8VDV9_MOUSE
ID Q8VDV9_MOUSE PRELIMINARY; PRT; 203 AA.
AC Q8VDV9;
```

```
DT      01-MAR-2002 (TREMBlrel. 20, Created)
DT      01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT      01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE 4932416A1IRik protein.
GN Name=Vwal; Synonyms=4932416A1IRik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX MEDLINE=22388257; PubMed=12479732; DOI=10.1073/pnae.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Wuzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.T., Skalski U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RA Strausberg R.L.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020136; AAH02136.1; -, mRNA.
DR Ensembl; ENSMUSG00000042115; Mus musculus.
DR MGI; MGI:2179729; 4932416A1IRik.
DR MGI; MGI:2179729; 4932416A1IRik.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS50853; FN3; 2.
DR Repeat.
SQ SEQUENCE 203 AA; 22068 MW; AAD7FDEF37626898 CRC64;
```

```
Query Match 36.3%; Score 782.5; DB 2; Length 203;
Best Local Similarity 73.3%; Pred. No. 1.2e-45;
Matches 151; Conservative 17; Mismatches 35; Indels 3; Gaps 1;

Qy      213 MRPQQLHATEITSSGFRIAMPPLLTADSGYVLELVPSAOPGAARQQLPGNATDWMAG 272
Db      1 MRPQQLHATEITSSGFRIAMPPLLTADSGYVLELVPSGKATTRQQLPGNATDWMAG 60
Qy      273 LDDPTDYVALVPSNVLLRPQLIRVTRPBEAGPERIVISHARPRSLRVSWAPALGSA 332
Db      61 LDDPTDYVALVPSNVLLRPQLIRVTRPBEAGPERIVISHARPRSLRVSWAPALGSA 120
Qy      333 AALGYHVOFGPLRGEAORVAVPAGRNCTTQGLAPGAYLVTVTAAPRSGRESALSAA 392
Db      121 AALGYHVOFGPLRGEAORVAVPAGRNCTTQGLAPGAYLVTVTAAPRSGRQRLSAA 180
Qy      393 CTPDGPDRPRPVPRAPPTGTASREP 418
Db      181 CTAAGATR---APGSMRPEAGPREP 203
```


CC		Isoid-P13944-1; Sequence=Displayed;
CC	Name=Short;	
CC	Isoid=P13944-2; Sequence=VSP_001148;	
CC	-1 TISSUE SPECIFICITY: Type XII collagen is present in tendons,	
CC	ligaments, pericardium, and periostrum, all dense connective	
CC	tissues containing type I collagen.	
CC	-1 DOMAIN: This sequence defines five distinct domains, two triple-	
CC	helical domains (COL1 and COL2) and three nontriple-helical	
CC	domains (NC1, NC2, and NC3).	
CC	-1 PWM: The triple-helical tail is stabilized by disulfide bonds at	
CC	each end.	
CC	-1 PWM: Prolines at the third position of the tripeptide repeating	
CC	unit (G-X-Y) are hydroxylated in some or all of the chains.	
CC	-1 PTM: O-glycosylated; glycosaminoglycan of chondrotin-sulfate type	
CC	(By similarity).	
CC	-1 SIMILARITY: Belongs to the fibril-associated collagens with	
CC	interrupted helices (FACIT) family.	
CC	-1 SIMILARITY: Contains 18 fibronectin type-II domains.	
CC	-1 SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.	
CC	-1 SIMILARITY: Contains 4 WFPA domains.	
CC	-----	
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use as long as its content is in no way modified and this statement is not	
CC	removed.	
CC	-----	
DR	EMBL; D00824; BAA00701.1; -; mRNA.	
DR	EMBL; X61024; CAA43358.1; -; mRNA.	
DR	EMBL; M17375; AAA48718.1; -; mRNA.	
DR	EMBL; J05137; AAA48635.1; -; mRNA.	
DR	EMBL; X67327; CAA47744.1; -; mRNA.	
DR	PIR; A40020; A40020.	
DR	HSSP; P56199; 1QC5.	
DR	Ensembl; ENSGALG00000015908; Gallus gallus.	
DR	InterPro; IPRO08160; Collagen.	
DR	InterPro; IPRO03961; FN_III.	
DR	InterPro; IPRO03129; Laminin_G_TSP_N.	
DR	InterPro; IPRO02035; WVF_A.	
DR	Pfam; PF001391; Collagen; 4.	
DR	Pfam; PF00041; Fn3; 17.	
DR	Pfam; PF00092; VWA; 4.	
DR	PRINTS; PR00453; VWFADOMAIN.	
DR	SMART; SM00060; FN3; 18.	
DR	SMART; SM00210; TSPN; 1.	
DR	SMART; SM00327; VWA; 4.	
DR	PROSITE; PS50853; FN3; 18.	
DR	PROSITE; PS50234; WVFA; 4.	
KW	Alternative splicing; Cell adhesion; Collagen;	
KW	Direct protein sequencing; Extracellular matrix; Glycoprotein;	
KW	Hydroxylation; Repeat; Signal; Structural protein.	
FT	SIGNAL	1 24 Potential.
FT	CHAIN	25 3124 Collagen alpha 1(XII) chain.
FT	DOMAIN	25 112 Fibronectin type-III 1.
FT	DOMAIN	139 311 WVFA 1.
FT	DOMAIN	332 421 Fibronectin type-III 2.
FT	DOMAIN	439 615 WVFA 2.
FT	DOMAIN	630 718 Fibronectin type-III 3.
FT	DOMAIN	721 809 Fibronectin type-III 4.
FT	DOMAIN	812 902 Fibronectin type-III 5.
FT	DOMAIN	905 993 Fibronectin type-III 6.
FT	DOMAIN	995 1083 Fibronectin type-III 7.
FT	DOMAIN	1086 1175 Fibronectin type-III 8.
FT	DOMAIN	1199 1371 WVFA 3.
FT	DOMAIN	1386 1472 Fibronectin type-III 9.
FT	DOMAIN	1474 1564 Fibronectin type-III 10.
FT	DOMAIN	1566 1654 Fibronectin type-III 11.
FT	DOMAIN	1655 1745 Fibronectin type-III 12.
FT	DOMAIN	1756 1845 Fibronectin type-III 13.
FT	DOMAIN	1847 1935 Fibronectin type-III 14.
FT	DOMAIN	1937 2026 Fibronectin type-III 15.
FT	DOMAIN	2028 2117 Fibronectin type-III 16.
FT	DOMAIN	2119 2206 Fibronectin type-III 17.

Query Match	20.3%: Score 437.5; DB 1; Length 3124;	
Best Local Similarity	31.1%: Pred. No. 1.2e-20;	
Matches 123; Conservative	66; Mismatches 166; Indels 39; Gaps 8	
FT DOMAIN 2210 2294	Fibronectin type-III 18.	
FT DOMAIN 2527 2500	WFPA 4.	
FT DOMAIN 2534 2716	TSP N-terminal.	
FT REGION 2455 2750	Nonhelical region (NC3).	
FT REGION 2751 2902	Triple-helical region (COL2) with 1 imperfection.	
FT REGION 2903 2945	Nonhelical region (NC2).	
FT REGION 2946 3048	Triple-helical region (COL1) with 2 imperfections.	
FT REGION 3049 3124	Nonhelical region (NC1).	
FT MOTIF 2899 2901	Cell attachment site (Potential).	
FT COMBRIAS 3086 3096	Asp/Glu-rich (acidic).	
FT COMBRIAS 3111 3123	Arg/Lys-rich (basic).	
FT CARBOHYD 32 32	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 797 797	O-linked (Xyl. . .) (chondroitin sulfate) (Potential).	
FT CARBOHYD 890 890	O-linked (Xyl. . .) (chondroitin sulfate) (Potential).	
FT CARBOHYD 981 981	O-linked (Xyl. . .) (chondroitin sulfate) (Potential).	
FT CARBOHYD 1006 1006	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 1032 1032	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 1044 1044	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 1512 1512	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 1767 1767	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 2210 2210	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 2273 2273	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 2532 2532	N-linked (GlcNAc. . .) (Potential).	
FT CARBOHYD 2633 2633	N-linked (GlcNAc. . .) (Potential).	
FT VANSPLIC 25 1188	Misling (in isoform Short).	
FT CONFLICT 1258 1258	/FTID=VSP_001148.	
FT CONFLICT 1264 1264	T -> S (in Ref. 4).	
FT CONFLICT 12759 2759	D -> E (in Ref. 4).	
FT CONFLICT 2803 2803	P -> A (in Ref. 2).	
FT CONFLICT 2977 2977	L -> F (in Ref. 2).	
FT CONFLICT 3075 3076	V -> F (in Ref. 2).	
FT CONFLICT 3124 AA; 340582 MM; 094285AFET7346CF CRC64;	QP -> AG (in Ref. 3).	
DB 32 RGDLMFLDSSASVSHYFSRVEFGQLVAPLPDGTGALBASLVHGSRPYTEPPFGQH 91		
DB 437 KADVFLVDGYSIGIANFVKRAFLVLSKFSFISPRKQISLVQSRDPHMEPSLAR 496		
QY 92 SSGEAAQAVAAASAKRMEDHTTGALVYAKQQLPRAAGARPQVVKLVMTDGGSSDPV 151		
DB 497 NRVDIIOAIWTFPYRGSTNTGKMTYRKEVFTSGSRPNPVRWILLITDGSSDAF 556		
QY 152 GPPQOLKDLGVWTFIVTSGRGNFLFLEAAASAPREKHLHV-DVDDLHIIVQELRGSI- 209		
DB 557 KEPAIKLADDAVEIFANGVAKAVRTELTAISPPETHVYVEBDARQIRISFELTQSV 616		
QY 210 -----LDAMR-----POOLHATEITSSGFRLLAMPPLLTADSGYVLELVSAOPGAR 257		
DB 617 LRIQELLAIRIKKSVVPAAKNVPSDVISDSKVSASAGSEKSLIRKKV--ALGGDEF 674		
QY 258 RQQLPQNAITMAGLDPTDIDVALVESNRLRIPQILNRKTRPEEAGPERIVISHAR 317		
DB 675 IVSVASSTSSVLNLLPETTYAAVSIAE-----YEDGQGPPLDGETTLEVGCA 724		
QY 318 PRSR-----VSWAPALGSAALGYNVQFQPLRGGEAQRVEVAGNCTTLOGLA 367		
DB 725 PRNRITDETDSFIVGWTPAPGN--VIRYRLVYKRLPLGGERRQVTSANERSTTLRLNI 782		
QY 368 PGTAIVLVTTAARSGRESALSAKACTPDGGRPRPR 403		
DB 783 PDTRYEVSVIAEYQSGPQNALNGYAKT--DEVKGNPR 817		


```
Q5VYK2_HUMAN
ID Q5VYK2_HUMAN PRELIMINARY; PRT; 2884 AA.
AC Q5VYK2;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Collagen, type XII, alpha 1.
GN Name=COL12A1; ORFNames=RP1-238D15.1-003;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OC NCBI_TaxID=9606;
RN NUCLEOTIDE SEQUENCE.
RP Tromans A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN NUCLEOTIDE SEQUENCE.
RP [2]
RN NUCLEOTIDE SEQUENCE.
RP Smith M.;
RN Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN NUCLEOTIDE SEQUENCE.
RP Corby N.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL354664; CAH71309.1; -; Genomic DNA.
DR EMBL; AL080250; CAI19897.1; -; Genomic DNA.
DR EMBL; AL096771; CAI19907.1; -; Genomic DNA.
DR EMBL; AL080250; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL096771; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL354664; CAI19897.1; JOINED; Genomic DNA.
DR EMBL; AL080250; CAI19907.1; JOINED; Genomic DNA.
DR EMBL; AL354664; CAI19907.1; JOINED; Genomic DNA.
DR EMBL; AL096771; CAI19897.1; JOINED; Genomic DNA.
DR Ensemble; ENSG0000011799; Homo sapiens.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0005578; Extracellular matrix (sensu Metazoa); IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00041; fn3; 18.
DR Pfam; PF00092; VWA; 4.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 18.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VWA; 4.
DR PROSITE; PS50234; FN3; 18.
DR PROSITE; PS50234; VWFA; 4.
DR Collagen; Extracellular matrix; Repeat; Structural protein.
SQ SEQUENCE 2884 AA; 315869 MW; 2D598F13656454D CRC64;

Query Match 19.7%; Score 424; DB 2; Length 2884;
Best Local Similarity 29.9%; Pred. 9.4e-20;
Matches 120; Conservative 68; Mismatches 186; Indels 26; Gaps 8;

QY 32 RQDLMLFLDLSASVSHYEFSSRYVEFGVGLVPLDGTGLRSLVHVGSRPYTFPPGQH 91
DB 438 KADIVFLVVGSSYSIGIANFVKVAFLEVLVKSFEISPNRVQISLVQYSRDPHTFLLKKF 497
QY 92 SSGEAAQDAVRSASQMGDTHTGLALVYAKEQLFAASGARPVPVKLVWVTDGSSDPV 151
DB 498 TKVEDIIEAINTFPYRGSGSTNTGKAMTYREKIFVSKSRSNVPMKMLITDGKSDAF 557
QY 152 GPMQELKGLGVTVFVSTGRGNFLSLAASAPAEKHLLFV-DVDDLHIIVQELGSI- 209
DB 558 RDPATKRSNDVEIFAVGVKDAVRSLEBALSPAEHTHYFTVDPFAQRISELVQSIC 617
QY 210 -----LDMR-----POOLHATEITSSGFRLLAMPPLLTADSGYVVLVPSAQPQGAAR 257
```

```
DB 618 LRIEQLAIKKKAVPPKDDSPFSSEVTSYGFRTNNSPGENVESYHI-----TYKEAG 671
QY 258 RQDL-----PGNATDWMAGLDPDPTDYDALVAVESNVRLLRPLQILRRTPREAGEERYIV 313
DB 672 DDEVTVVPEASTSVLLSLKRETIYLVNVAEYDGRSIFLAGETIEVGVGARNLKV 731
QY 314 SHARPRLSVSWAPALGSAALGVYVQFGLRGSGAQRVEVPAGNCTTLOGLAFGTAYL 373
DB 722 TDETTDSFKITWTQAPG--RVLRVRIIVRPVAGSGESREVTTPPNORRTLEMLIDTKYE 789
QY 374 VVTVAAPRSGRSALSACACTPDGRRPRPRVPRAPRTGTAS 415
DB 790 VSVIPEYFSGPGLTGNAAITE-VRGNPRDL-RVSDPTTST 829
```

RESULT 12

Q5VYK1_HUMAN

ID Q5VYK1_HUMAN PRELIMINARY; PRT; 3063 AA.

AC Q5VYK1;

DT 01-FEB-2005 (TrEMBLrel. 29, Created)

DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Collagen, type XII, alpha 1

GN Name=COL12A1; ORFNames=RP1-238D15.1-001;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;

OC Homo.

OC NCBI_TaxID=9606;

RN NUCLEOTIDE SEQUENCE.

RP [1]

RN NUCLEOTIDE SEQUENCE.

RP Tromans A.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

RN NUCLEOTIDE SEQUENCE.

RP Smith M.;

RN Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

RN NUCLEOTIDE SEQUENCE.

RP Corby N.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL354664; CAH71310.1; -; Genomic DNA.

DR EMBL; AL080250; CAI19898.1; -; Genomic DNA.

DR EMBL; AL096771; CAI19908.1; -; Genomic DNA.

DR EMBL; AL080250; CAH71310.1; JOINED; Genomic DNA.

DR EMBL; AL096771; CAH71310.1; JOINED; Genomic DNA.

DR EMBL; AL354664; CAI19898.1; JOINED; Genomic DNA.

DR EMBL; AL080250; CAI19908.1; JOINED; Genomic DNA.

DR EMBL; AL354664; CAI19908.1; JOINED; Genomic DNA.

DR EMBL; AL096771; CAI19898.1; JOINED; Genomic DNA.

DR Ensemble; ENSG0000011799; Homo sapiens.

DR GO; GO:0005737; Cytoplasm; IEA.

DR GO; GO:0005578; Extracellular matrix (sensu Metazoa); IEA.

DR GO; GO:0005198; P:structural molecule activity; IEA.

DR GO; GO:0007155; P:cell adhesion; IEA.

DR GO; GO:006817; P:phosphate transport; IEA.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR003961; FN_III.

DR InterPro; IPR003129; Laminin_G_TSP_N.

DR InterPro; IPR002035; VWF_A.

DR Pfam; PF01391; Collagen; 4.

DR Pfam; PF00041; fn3; 18.

DR Pfam; PF00092; VWA; 4.

DR PRINTS; PR00453; VWFADOMAIN.

DR SMART; SM00060; FN3; 18.

DR SMART; SM00210; TSPN; 1.

DR SMART; SM00327; VWA; 4.

DR PROSITE; PS50234; FN3; 18.

DR PROSITE; PS50234; VWFA; 4.

DR Collagen; Extracellular matrix; Repeat; Structural protein.

SQ SEQUENCE 3063 AA; 333147 MW; EA38CAFECB893D2 CRC64;

Query Match 19.7%; Score 424; DB 2; Length 3063;
Best Local Similarity 29.9%; Pred. No. 1e-15;
Matches 120; Conservative 68; Mismatches 186; Indels 28; Gaps 8;

32 RGDLMFLDSSASVSHYEFSEFVGOVAPLPGTGLRSLVHVSRRPTEFPQGH 91
438 KADIVFLVDGSIIGIANFVKRAFLVLYKSFELSPNRVQSLVQSRDPTEETLKKF 497
92 SSGEAAQDAVRAAQRMDPTHTLALVYAKQLFAEASGAPGVKVLVWTDGSSDPV 151
498 TKVEDIIRAINTFPRYGSGTNGKAMTYREKIFVPSKSGSRGNVKKMILLIDGSSDAF 557
152 GPPMELDLQGTIVTVISGRGNFELSAASAAPAEKHHPV-DVDDHLITVQELRGSI- 209
558 RDPALIKNSVDIEIFAVGVKQAVRSELEAIASPPAEITHVFTVEDDAFORISFELTQSI 617
210 -----LDAMR-----POOLATEITSSGFFLAMPPLLTADSGYVLELVSAQGAAR 257
618 LRIBELAIKKKAVVPPDLSFSEVTSYGFKTMSPAGEVNFSTHI-----TYKEANG 671
258 RQQL-----PGNATDWMAGLDPTDYVALVESNVRLRPOOLRVRTRPEAGERIATV 313
672 DDEVTVPEASTSVLSLKPETLYLVNVTAEYEDGFSIPLAGETTEEYVGADRNLIKV 731
314 SHARRSLRVSWAPALGSAALGYVQFPLRGSAQRVYEVAGNCTTLOGLAGTAYL 373
732 TDERTDSKRTWTQAPG--RVLRYRIYRPVAGSGSRREVTTPPNRRRLTENVLIDTYKE 789
374 VVTVAFRSGRESALSAKACTPDGPRPRPVPRAPTPTAS 415
790 VSVIPEYFGPGCTPLTGNABTEE-VKGNPRDL-RVSDPTTST 829

RESULT 13
COCAL HUMAN
ID COCAL HUMAN STANDARD; PRT; 3063 AA.
AC Q99715; Q99715; (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN Name=COL12A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT), AND PARTIAL PROTEIN
RA MEDLINE=97288521; PubMed=9143499; DOI=10.1006/geno.1997.4638;
RA Gerecke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,
Hudson D.L., Champilaud M.-F., Olsen B.R., Burgeson R.E.,
"Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(IX)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human
RT chromosome 6q12-q13.";
RT Genomics 41:236-242(1997).
RL
CC -!- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the perifibrillar matrix (By similarity).
CC -!- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment: The final tissue form of collagen XII may contain
CC homotrimers of either isoform long or isoform short or any
CC combination of isoform long and isoform short;
CC Name=Long;
CC IsoId=Q99715-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q99715-2; Sequence=VSP_001149;

CC -!- TISSUE SPECIFICITY: Found in collagen I-containing tissues: both
CC isoform short and isoform long appear in amnion, chorion, skeletal
CC muscle, small intestine, and in cell culture of dermal
CC fibroblasts, keratinocytes and endothelial cells. Only the short
CC isoform is found in lung, placenta, kidney and a squamous cell
CC carcinoma cell line.
CC -!- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end (By similarity).
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains (By
CC similarity).
CC -!- PTM: O-glycosylation of isoform long; glycosaminoglycan of
CC chondroitin-sulfate type (By similarity).
CC -!- SIMILARITY: Belongs to the fibril-associated collagens with
CC interrupted helices (FACIT) family.
CC -!- SIMILARITY: Contains 18 fibronectin type-III domains.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -!- SIMILARITY: Contains 4 VMPA domains.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U73778; AAC51244.1; -; mRNA.
CC EMBL; U73779; AAD40483.1; -; mRNA.
CC HSSP; P18614; IMHP.
CC DR Ensembl; ENSG00000111799; Homo sapiens.
CC HGNC; HGNC:2188; COL12A1.
CC MIM; 120320; -;
CC GO; GO:0005595; C:collagen type XII; TAS.
CC GO; GO:0001501; P:skeletal development; TAS.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR002035; VWF_A.
CC Pfam; PF01391; Collagen; 4.
CC Pfam; PF00041; Fn3; 18.
CC DR Pfam; PF00092; VWA; 4.
CC DR PRINTS; PR00453; VMPADOMAIN.
CC DR PROSITE; PS50853; FN3; 18.
CC DR PROSITE; PS50234; VWFA; 4.
CC KW Alternative splicing; Cell adhesion; Collagen;
CC direct protein sequencing; Extracellular matrix; Glycoprotein;
CC Hydroxylation; Repeat; Signal; Structural protein.
CC KW
CC SIGNAL 1 24
CC CHAIN 25 3063 Collagen alpha 1(XII) chain.
CC DOMAIN 25 112 Fibronectin type-III 1.
CC DOMAIN 140 316 VMPA 1.
CC DOMAIN 333 422 Fibronectin type-III 2.
CC DOMAIN 440 616 VMPA 2.
CC DOMAIN 631 719 Fibronectin type-III 3.
CC DOMAIN 722 810 Fibronectin type-III 4.
CC DOMAIN 813 901 Fibronectin type-III 5.
CC DOMAIN 904 993 Fibronectin type-III 6.
CC DOMAIN 995 1083 Fibronectin type-III 7.
CC DOMAIN 1086 1175 Fibronectin type-III 8.
CC DOMAIN 1199 1371 VMPA 3.
CC DOMAIN 1384 1472 Fibronectin type-III 9.
CC DOMAIN 1474 1563 Fibronectin type-III 10.
CC DOMAIN 1565 1652 Fibronectin type-III 11.
CC DOMAIN 1654 1743 Fibronectin type-III 12.
CC DOMAIN 1752 1841 Fibronectin type-III 13.
CC DOMAIN 1843 1931 Fibronectin type-III 14.
CC DOMAIN 1933 2022 Fibronectin type-III 15.
CC DOMAIN 2024 2113 Fibronectin type-III 16.
CC DOMAIN 2115 2202 Fibronectin type-III 17.
CC DOMAIN 2206 2290 Fibronectin type-III 18.
CC DOMAIN 2292 2333 VMPA 4.
CC DOMAIN 2333 2496 TSP N-terminal.
CC DOMAIN 2520 2712 Nonhelical region (NC3).
CC DOMAIN 2451 2746 Triple-helical region (COL2) with 1
CC REGION 2747 2898 imperfection.
CC FT

FT	REGION	2899	2941	Nonhelical region (NC2).
FT	REGION	2942	3044	Triple-helical region (COL1) with 2 imperfections.
FT	REGION	3045	3063	Nonhelical region (NC1).
FT	MOTIF	862	864	Cell attachment site (Potential).
FT	MOTIF	2779	2781	Cell attachment site (Potential).
FT	MOTIF	2895	2897	Cell attachment site (Potential).
FT	MOD_RES	2944	2944	Hydroxyproline (By similarity).
FT	MOD_RES	2947	2947	Hydroxyproline (By similarity).
FT	MOD_RES	2950	2950	Hydroxyproline (By similarity).
FT	MOD_RES	2959	2959	Hydroxyproline (By similarity).
FT	MOD_RES	2965	2965	Hydroxyproline (By similarity).
FT	MOD_RES	2968	2968	Hydroxyproline (By similarity).
FT	MOD_RES	2971	2971	Hydroxyproline (By similarity).
FT	MOD_RES	2983	2983	Hydroxyproline (By similarity).
FT	MOD_RES	3000	3000	Hydroxyproline (By similarity).
FT	MOD_RES	3003	3003	Hydroxyproline (By similarity).
FT	MOD_RES	3014	3014	Hydroxyproline (By similarity).
FT	MOD_RES	3023	3023	Hydroxyproline (By similarity).
FT	MOD_RES	3026	3026	Hydroxyproline (By similarity).
FT	MOD_RES	3029	3029	Hydroxyproline (By similarity).
FT	CARBOHYD	700	700	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	798	798	O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT	CARBOHYD	889	889	O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT	CARBOHYD	981	981	O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT	CARBOHYD	1763	1763	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	2206	2206	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	2528	2528	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	2679	2679	N-linked (GlcNAc...) (Potential).
FT	VARSPPLIC	25	1188	Missing (in isoform Short).
FT	SEQUENCE	3063 AA;	333194 MW;	75FEA78FA8E8293 CRC64;
Query Match		19.6%;	Score 421.5;	DB 1; Length 3063;
Best Local Similarity		27.5%;	Pred. No. 1.5e-19;	
Matches 128;		Conservative 59;	Mismatches 174;	Indels 105; Gaps 9;
QY	32	RDLMFLDSSASVSHYFSRVRFEVGVLPGLTGALRASLVHVGSRPTPEFGH	91	
DB	438	KADIVFLVDGYSIGIANFVKARAFLEVLVKSFEISPNVQISLVYSDPHTEFLTKF	497	
QY	92	SSGEAQAQAVRASQOMGDTHGTALVYAKQQLFAEASGARQGVPLVWMTDGGSSDPV	151	
DB	498	TKVEDIIEIINFPYRGGSTNTGKATYREKIFVPSKSRNVKVMILLIDGKSSDAF	557	
QY	152	GPOMELKDLGTVFIVSTGRGNFLELSAASAPAEKHLFV-DVDDLHIYQELRGS1-	209	
DB	556	RDPALIKRNSDVEIFAVGVKDAVRSLELAISPAPATHVFTVBDPAFORISFELQSIG	617	
QY	210	-----LDAMR-----POOLHATEITSSGFRLANPPL-----LTAOSGYV	244	
DB	618	IRIEQELAIKKKAAVVPKDLSEFSEVTSYGFYTNMSPAGEVNSYHITKEAAGDEEVV	677	
QY	245	LE-----LVPSAQP-----GAARQOLPENATD	267	
DB	678	VAPASTSVLSSLKPEETLYLVNVTAAEYDGSIFLAGETTEBYKAGARINKLVDEDTTD	737	
QY	268	-----WIVA-----GLDPTDYDVALVPSN	288	
DB	738	SFKITWTQAPGRVLCRIIYRPVAGSESEVTPPNQRRRTLENLIPDKYEVSIVPEVF	797	
QY	289	VLLRPQILIRVTRPREAGPERIVISHAPRSIRVSWAPALGSAALGYHVOFGPLRGE	348	
DB	798	SGPGTPTGNAATEEVRGNPRDLRVSDPTSTKLSWGAPEGVVKO--LVVYTTVPAQGE	855	
QY	349	AORVEVPAGRNCTLOGLAPGTAIVTVTAAPRSGESALSAKAC	394	
DB	856	TOEVTVRGDTTIVTLOGLEKGTQYALSVTALYASGADALFGEGET	901	

RESULT 14	Q4SD22_TETNG	PRT;	1723 AA.
AC	Q4SD22_TETNG		
DT	13-SEP-2005 (TREMBlrel. 31, Created)		
DT	13-SEP-2005 (TREMBlrel. 31, Last sequence update)		
DT	13-SEP-2005 (TREMBlrel. 31, Last annotation update)		
DE	Chromosome 14 SCAF14645, whole genome shotgun sequence.		
GN	ORFNames=GSTENG00020231001;		
OS	Tetradodon nigroviridis (Green puffer).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;		
OC	Tetraodontidae; Tetraodontidae; Tetraodon.		
OX	NCBI_TaxID=99883;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Tailon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,		
RA	Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,		
RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,		
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,		
RA	Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,		
RA	Biemont C., Shaili Z., Catrolicco L., Poulain J., De Berardinis V.,		
RA	Cruaud C., Duprat S., Broctier P., Coutanceau J.P., Guzy J.,		
RA	Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,		
RA	Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,		
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,		
RA	Winkler P., Lander E.S., Weissbach J., Roset Crollins H.,		
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals		
RL	the early vertebrate proto-karyotype."		
RL	Nature 431:946-957(2004).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RG	Genoscope, Whitehead Institute Centre for Genome Research,		
RL	Submitted (FBS-2004) to the EMBL/Genbank/DBS databases.		
CC	-1- CAUTION: The sequence shown here is derived from an		
CC	EMBL/Genbank/DBS whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; CA001014645; CAG01460.1; -; Genomic_DNA.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR003962; FNIII_subd.		
DR	InterPro; IPR02035; WVF_A.		
DR	Pfam; PF00041; fn3; 11.		
DR	Pfam; PF00092; WMA; 3.		
DR	PRINTS; PR00014; ENTYPBIII.		
DR	PRINTS; PR00453; WVFADOMAIN.		
DR	SMART; SM00060; FN3; 11.		
DR	SMART; SM00327; WMA; 3.		
DR	PROSITE; PS00853; FN3; 12.		
KW	PROSITE; PS00334; WMA; 3.		
KW	Collagen; Extracellular matrix; Structural protein.		
FT	NON TER		
FT	NON TER		
FT	NON TER		
SEQUENCE	1723 AA;	187496 MW;	7EB78DA8D1E6B40 CRC64;
Query Match		19.1%;	Score 411.5; DB 2; Length 1723;
Best Local Similarity		26.5%;	Pred. No. 3.6e-19;
Matches 128;		Conservative 54;	Mismatches 180; Indels 121; Gaps 8;
QY	32	RDLMFLDSSASVSHYFSRVRFEVGVLPGLTGALRASLVHVGSRPTPEFGH	91
DB	436	QADVVLVDGYSIGILANPAKARALEVLVNFDFDGPDKVQISLVYSDPHTEFLYLDH	495
QY	92	SSGEAQAQAVRASQOMGDTHGTALVYAKQQLFAEASGARQGVPLVWMTDGGSSDPV	151
DB	496	HNLEAVLVALTRFPYRGGSTNTGRAMTVRETVFOASRGARAVRVTILLIDGKSSDAF	555
QY	152	GPOMELKDLGTVFIVSTGRGNFLELSAASAPAEKH-----	189
DB	556	QEPANLRNSDVEIFAVGVKDAVRSLELAISPAPATHVYTVEDDPAFORISTELTQSIG	615

QY	190	-----LHVYV-----	-----DLHIT	201
Db	616	LRIEQLQINQRRLVQPRDLVFSVDGPPRSFRASWEINANNVESYIVQRPTEGEBSHYV	-----	675
QY	202	VGELGSIIDARP	-----OOLHATEIT	224
Db	676	SMSVEDVLTALLPHILPTLRVSVSAQYAKTSLPTVGYGTTAEERGSVOMLKVTEES	-----	735
QY	225	SSGFRILAMPPLITADSGYVVLIELVPSAOPGAARQOL	-----PGNATDWMAGIDPPTDYD	280
Db	736	PGSFRVSWAAGAVARYRL-----TYQPAAGAEQLEAFTAGPELMTVLVQDLOPRTTYR	-----	790
QY	261	VALVPSNRLLRPQILRVTRPEEAQPERIVISHPRNSLRVSWAPALGSAALGYHVO	-----	340
Db	791	VVLTPEYKKGPGGPGQDTGDTTEBAKSPGGLRVFDQTVSMKYSWEPAPD--VLQYRLA	-----	848
QY	341	PGPLGGEAGQREVPAGRNCTTLOGLAPGAVLYVTYTAAPRSGRSEALSAAKATPD--GP	-----	398
Db	849	YRASSGGPKEVSVKCTNTAIVLKKQDGTETDITVSARYRSLGELKLGKATLEEVGP	-----	908
QY	399	RPR 401		
Db	909	-PR 910		
RESULT 15				
COCAL MOUSE				
ID	COCAL MOUSE	STANDARD;	PRT;	3119 AA.
AC	O60847: P70322;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	Collagen alpha 1(XII) chain precursor.			
GN	Name=Coll2a1;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;			
OC	Muroidea; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XI1A-1 AND XI1B-1).			
RC	STRAIN=C57BL/6J, and Swiss Webster; TISSUE=Skin;			
RX	MEDLINE=96170761; PubMed=8601036;			
RA	Boehme K., Li Y., Oh P.S., Olsen B.R.;			
RT	"Primary structure of the long and short splice variants of mouse collagen XII and their tissue-specific expression during embryonic development.";			
RL	dev. Dyn. 204:432-445(1995).			
RP	[2]			
RP	PARTIAL NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XI1A-2 AND XI1B-2).			
RC	STRAIN=C57BL/6J; TISSUE=Skin fibroblast;			
RX	MEDLINE=99348349; PubMed=10419532; DOI=10.1074/jbc.274.31.22053;			
RA	Kania A.M., Nishchenberger E., Baur S.T., Karimub N.Y., Taylor R.W.,			
RT	Olsen B.R., Mishumura I.;			
RT	"Structural variation of type XII collagen at its carboxyl-terminal NCI domain generated by tissue-specific alternative splicing.";			
RL	J. Biol. Chem. 274:22053-22059(1999).			
CC	-1- FUNCTION: Type XII collagen interacts with type I collagen-			
CC	containing fibrils, the COL1 domain could be associated with the			
CC	surface of the fibrils, and the COL2 and NC3 domains may be			
CC	localised in the perifibrillar matrix (By similarity).			
CC	-1- SUBUNIT: Trimer of identical chains each containing 190 kDa of			
CC	nontriple-helical sequences (By similarity).			
CC	-1- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=4;			
CC	Comment=The final tissue form of collagen XII may contain			
CC	homocimers or any combination of the various isoforms;			
CC	Name=XI1A-1;			
CC	Isoid=Q60847-1; Sequence=Displayed;			
CC	Name=XI1A-2; Synonyms=ER#K;			
CC	Isoid=Q60847-2; Sequence=VSP 001151, VSP 001152;			

CC	Name=XIIB-1;
CC	Isoid=G60847-3; Sequence=VSP_001150;
CC	Name=XIIB-2;
CC	Isoid=G60847-4; Sequence=VSP_001150, VSP_001151, VSP_001152;
CC	-1-TISSUE SPECIFICITY: Highest expression in tendons, perichondrium,
CC	skin, cornea, sclera, blood vessels, and periosteum.
CC	-1-DEVELOPMENTAL STAGE: The long NC3 XIIA isoforms are predominant at
CC	early stages (ED7 and 11); at later stages of development (ED15
CC	and 17) the short NC3 XIIB forms become the major forms. As the
CC	short NC3 forms become the major product, the long splice variant
CC	continues to be expressed in several tissues, even after birth.
CC	The long NC1 isoforms, XIIA-1 and XIIB-1, peak in 15-day old
CC	embryos and decrease in 17-day old ones. The expression of the
CC	short NC1 form XIIB-2 remains constant throughout late stages of
CC	embryonic development (ED15 and ED17).
CC	-1-PMW: The triple-helical tail is stabilized by disulfide bonds at
CC	each end (By similarity).
CC	-1-PWM: Proline at the third position of the tripeptide repeating
CC	unit (G-X-Y) are hydroxylated in some or all of the chains (By
CC	similarity).
CC	-1-PWM: O-glycosylation of isoform XIIA-2; glycosaminoglycan of
CC	chondroitin-sulfate type (By similarity).
CC	-1-SIMILARITY: Belongs to the fibril-associated collagens with
CC	interrupted helices (FACIT) family.
CC	-1-SIMILARITY: Contains 18 fibronectin type-III domains.
CC	-1-SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC	-1-SIMILARITY: Contains 4 VWFA domains.
CC	-----
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use as long as its content is in no way modified and this statement is not
CC	removed.
CC	-----
DR	EMBL; U25652; AAA9719.1; ALT_SEQ; mRNA.
DR	EMBL; U57095; AAB07047.1; -; mRNA.
DR	HSSP; F18614; IMHP.
DR	Ensembl; ENSMUSG00000032332; Mus musculus.
DR	MGI; MGI:88448; Coll2a1.
DR	GO; GO:0005615; C:extracellular space; TAS.
DR	InterPro; IPR008160; Collagen.
DR	InterPro; IPR003961; FN_III.
DR	InterPro; IPR003129; Laminin_G_TSP_N.
DR	InterPro; IPR002035; VWF_A.
DR	Pfam; PF01391; Collagen; 4.
DR	Pfam; PF00041; Fn3; 18.
DR	Pfam; PF00092; VWA; 4.
DR	PRINTS; PR00453; VWFPADOMAIN.
DR	SMART; SMO0060; FN3; 18.
DR	SMART; SMO0210; TSPN; 1.
DR	SMART; SMO0327; VMA; 4.
DR	PROSITE; PS50853; FN3; 18.
DR	PROSITE; PS50234; VWFA; 4.
KW	Alternative splicing; Cell adhesion; Collagen; Extracellular matrix;
KW	Glycoprotein; Hydroxylation; Repeat; Signal; Structural protein.
FT	SIGNAL 1 24 Potential.
FT	CHAIN 25 319 Collagen alpha 1(XII) chain.
FT	DOMAIN 25 112 Fibronectin type-III 1.
FT	DOMAIN 140 316 VWFA 1.
FT	DOMAIN 333 422 Fibronectin type-III 2.
FT	DOMAIN 444 620 VWFA 2.
FT	DOMAIN 635 723 Fibronectin type-III 3.
FT	DOMAIN 726 814 Fibronectin type-III 4.
FT	DOMAIN 817 905 Fibronectin type-III 5.
FT	DOMAIN 908 997 Fibronectin type-III 6.
FT	DOMAIN 999 1087 Fibronectin type-III 7.
FT	DOMAIN 1090 1179 Fibronectin type-III 8.
FT	DOMAIN 1203 1375 VWFA 3.
FT	DOMAIN 1388 1476 Fibronectin type-III 9.
FT	DOMAIN 1478 1567 Fibronectin type-III 10.
FT	DOMAIN 1569 1656 Fibronectin type-III 11.
FT	DOMAIN 1660 1747 Fibronectin type-III 12.
FT	DOMAIN 1758 1847 Fibronectin type-III 13.

```

FT DOMAIN 1849 1937 Fibronectin type-III 14.
FT DOMAIN 1939 2028 Fibronectin type-III 15.
FT DOMAIN 2030 2119 Fibronectin type-III 16.
FT DOMAIN 2121 2208 Fibronectin type-III 17.
FT DOMAIN 2212 2298 Fibronectin type-III 18.
FT DOMAIN 2329 2501 vWFA 4.
FT DOMAIN 2525 2717 TSP N-terminal.
FT REGION 2525 2751 Nonhelical region (NC3).
FT REGION 2752 2899 Triple-helical region (COL2) with 1
FT REGION 2900 2942 Nonhelical region (NC2).
FT REGION 2943 3045 Triple-helical region (COL1) with 2
FT REGION 3046 3119 Imperfections.
FT MOTIF 3046 3119 Nonhelical region (NC1).
FT MOTIF 3119 Cell attachment site (Potential).
FT MOTIF 3119 Cell attachment site (Potential).
FT MOTIF 3119 Cell attachment site (Potential).
FT COMBIAS 2896 2898 Poly-Thr.
FT MOD_RES 2896 2898 Hydroxyproline (By similarity).
FT MOD_RES 2945 2945 Hydroxyproline (By similarity).
FT MOD_RES 2948 2948 Hydroxyproline (By similarity).
FT MOD_RES 2951 2951 Hydroxyproline (By similarity).
FT MOD_RES 2960 2960 Hydroxyproline (By similarity).
FT MOD_RES 2966 2966 Hydroxyproline (By similarity).
FT MOD_RES 2969 2969 Hydroxyproline (By similarity).
FT MOD_RES 2972 2972 Hydroxyproline (By similarity).
FT MOD_RES 2984 2984 Hydroxyproline (By similarity).
FT MOD_RES 3001 3001 Hydroxyproline (By similarity).
FT MOD_RES 3004 3004 Hydroxyproline (By similarity).
FT MOD_RES 3015 3015 Hydroxyproline (By similarity).
FT MOD_RES 3024 3024 Hydroxyproline (By similarity).
FT MOD_RES 3027 3027 Hydroxyproline (By similarity).
FT MOD_RES 3030 3030 Hydroxyproline (By similarity).
FT CARBOHYD 802 802 O-linked (GlcNAc...) (Potential).
FT CARBOHYD 893 893 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 985 985 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 1769 1769 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2212 2212 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2533 2533 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2684 2684 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 25 1190 Missing (in isoform XIIB-1 and isoform
FT VARSPLIC 3062 3064 /FTId=VSP_001150.
FT VARSPLIC 3065 3119 EPR -> GSG (in isoform XIIB-2 and isoform
FT VARSPLIC 3065 3119 /FTId=VSP_001151.
FT VARSPLIC 3065 3119 Missing (in isoform XIIB-2 and isoform
FT VARSPLIC 3065 3119 XIIB-2).
SQ SEQUENCE 3119 AA; 340244 MM; 981F95C86AB3251 CRC64;
Query Match 18.9%; Score 407; DB 1; Length 3119;
Best Local Similarity 28.7%; Pred. No. 1.5e-18;
Matches 121; Conservative 64; Mismatches 177; Indels 60; Gaps 10;

```

```

QY 236 -----LTPDSGYVLELVPSAQPGAAARQQLPGNATDWTMGADPDPTDYDVA 282
DB 660 GDNVESHYTYXDANGDDEVTVE-----PASTSVLMSLRRETLTYLVN 704
QY 283 LVPESVRLRLRPQILRVTRPREAG-PERIVISHARPRSLRYSWAPALGSAALGVHOF 341
DB 705 VPAEVEDGSPV-ITGEETTAELVKGVPNNLKVTDETTDSFKLTWSQAPG-RVLYRIRIX 761
QY 342 GPLRGGEAQRVWPVAGRNCTTLQGLAPGYAVLVTVTAAPRSGRESALSAKACTPDGPRPR 401
DB 762 RVPSSGESKEVSTPANQRKTELENLTPDKIEISVIAEYPSGPGSPLTGNAAEE-VRGN 820
QY 402 PR 403
DB 821 PR 822

```

Search completed: February 13, 2006, 07:47:47
Job time : 331.926 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:33:34 ; Search time 254.435 Seconds
(without alignments)
721.837 Million cell updates/sec

Title: US-10-699-035A-6

Perfect score: 2154
Sequence: 1 MLPTALGLALSLRLALARS.....RRPRPVPRAPPTGRASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq_21.*
2: geneseqp1908.*
3: geneseqp1908.*
4: geneseqp2000.*
5: geneseqp2000.*
6: geneseqp2003as.*
7: geneseqp2003as.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2154	100.0	418	6 AAE32502	Aae32502 Human w11
2	2130.5	98.9	445	5 ABP69674	Abp69674 Human pol
3	2130.5	98.9	445	8 ADH71106	Adh71106 Human pro
4	1653.5	76.8	415	6 AAE32501	Aae32501 Mouse w11
5	1643.5	76.3	421	6 AAE6326	Aae6326 Kidney in
6	1362	63.2	299	3 AAB42581	Aab42581 Human ORF
7	1288.5	59.8	285	4 AAB88340	Aab88340 Human mem
8	1288.5	59.8	285	9 ADY63045	Ady63045 Human clo
9	1084	50.3	215	4 AAB87344	Aab87344 Human gen
10	1084	50.3	215	5 AAB65347	Abg65347 Human alb
11	1084	50.3	215	8 ADL78614	Adl78614 Humanin f
12	1084	50.3	242	4 AAB87418	Aab87418 Human gen
13	1077	50.0	226	4 AAB87424	Aab87424 Human gen
14	918	42.6	185	4 AAE03654	Aae03654 Human ext
15	918	42.4	180	6 AAE32500	Aae32500 Human von
16	768	35.7	180	6 AAE32503	Aae32503 Mouse von
17	424	19.7	3063	6 ABR47416	Abt47416 Breast ca
18	424	19.7	3063	8 ADJ75666	Adj75666 Marker ge
19	421.5	19.6	3063	5 ABB90762	Abb90762 Human Tum
20	421.5	19.6	3063	6 ABU54469	Abu54469 Human tum
21	421.5	19.6	3063	6 ABR47415	Abt47415 Breast ca
22	421.5	19.6	3118	8 ADU27790	Adu27790 Human ful
23	407	18.9	3067	8 ADJ76366	Adj76366 Marker ge
24	392.5	18.2	1207	7 ADJ69785	Adj69785 Human hea

25	392.5	18.2	1279	8	ADS98011	Ads98011 Protein f
26	392.5	18.2	1297	7	ADM47269	Adm47269 Fn3/TFPN/
27	392.5	18.2	1297	8	ADS98013	Ads98013 Protein f
28	392.5	18.2	1329	8	ADS98012	Ads98012 Protein f
29	384	17.8	77	3	AAG02196	Aag02196 Human sec
30	382.5	17.8	2502	8	ADQ39964	Adq39964 Human myo
31	382.5	17.8	2944	6	ABH56436	Abh56436 Human can
32	382.5	17.8	2944	8	ADQ18966	Adq18966 Human sof
33	382.5	17.8	2944	8	ADQ39966	Adq39966 Human myo
34	382.5	17.8	2944	8	ADQ06300	Adq06300 Novel bro
35	382.5	17.8	2944	9	ADK05807	Adk05807 Cyclin-de
36	360	16.7	1899	8	ADJ75667	Adj75667 Marker ge
37	326	15.1	1780	7	ADP15980	Adp15980 G-coupled
38	326	15.1	1780	8	ADL93919	Adl93919 Human G-c
39	284.5	13.2	401	8	ADP98648	Adp98648 Protein f
40	272.5	12.7	176	5	ABH78816	Abh78816 Von Wille
41	270	12.5	795	4	AAB27229	Aab27229 Human EXM
42	268.5	12.5	644	8	ADR66963	Adr66963 Mouse can
43	268.5	12.5	644	9	ADJ25632	Adj25632 Murine ca
44	268.5	12.5	843	7	ADJ68211	Adj68211 Human hea
45	264	12.3	755	5	AAU84267	Aau84267 Human end

ALIGNMENTS

RESULT 1	AAE32502	standard; protein; 418 AA.
XX	AAE32502;	
XX	24-MAR-2003	(first entry)
XX	Human Willebrand Factor A domain related-protein (WARP).	
XX	Human Willebrand Factor A domain related-protein (WARP).	
XX	Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;	
XX	extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;	
XX	gene therapy; human.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
XX	Peptide	1..18
XX	Protein	/label= Signal_peptide
XX	Modified-site	/note= "Human mature WARP protein"
XX	Misc-difference	/note= "O-glycosylation site"
XX	Modified-site	/note= "Encoded by CTCGCG"
XX	Modified-site	/note= "N-glycosylation site"
XX	Modified-site	/note= "N-glycosylation site"
XX	Modified-site	/note= "N-glycosylation site"
XX	Disulfide-bond	/note= "O-glycosylation site"
XX	WO200208184-A1.	
XX	07-NOV-2002.	
XX	02-MAY-2002; 2002WO-AU000542.	
XX	02-MAY-2001; 2001AU-00004701.	
XX	(MURD-) MURDOCH CHILDRENS RES INST.	
XX	Bateman JF, Fitzgerald DJ;	
XX	WPI; 2003-111873/10.	
XX	N-PSDB; AAD50399.	

XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
PT the manufacture of a medicament in the treatment of a disease condition
PT of the extracellular matrix, in particular arthritis.

XX Claim 11; Page 76-78; 103pp; English.

XX The invention relates to Willebrand Factor A domain related-protein
CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a
CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medicament in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
CC present sequence is human WARP protein

XX Sequence 418 AA;

Query Match 100.0%; Score 2154; DB 6; Length 418;
Best Local Similarity 100.0%; Pred. No. 2.6e-179;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLPTALGLALSLRLALRSGAERGPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
DB 1 MLPTALGLALSLRLALRSGAERGPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
QY 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVARASQRMGDHTTGALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVARASQRMGDHTTGALVYA 120
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLEISA 180
DB 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLEISA 180
QY 181 AASAPAEKHLHFVDVDDHLIIYQELRGSLIDAMRPOQLHATEITSSGFLAMPPLLTADS 240
DB 181 AASAPAEKHLHFVDVDDHLIIYQELRGSLIDAMRPOQLHATEITSSGFLAMPPLLTADS 240
QY 241 GYVLELVPSAOPGAARQOPLGNATDWTWAGLDPTDYDALVYESNRLRLPOLILVR 300
DB 241 GYVLELVPSAOPGAARQOPLGNATDWTWAGLDPTDYDALVYESNRLRLPOLILVR 300
QY 301 TPBEAGPERIVISHARPSRLVSNWAPALGSAALGYHGFQPLRGGEAQRVEVPAGRNC 360
DB 301 TPBEAGPERIVISHARPSRLVSNWAPALGSAALGYHGFQPLRGGEAQRVEVPAGRNC 360
QY 361 TTLQGLAPGTALVLTVAFAFRSGRESALSAKACTPDGPRPRRPVPRAPTGTASREP 418
DB 361 TTLQGLAPGTALVLTVAFAFRSGRESALSAKACTPDGPRPRRPVPRAPTGTASREP 418

```

RESULT 2
ABP69674
ID ABP69674 standard; protein; 445 AA.

XX ABP69674;

XX 20-JAN-2003 (first entry)

XX Human polypeptide SEQ ID NO 1721.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
XX cell-proliferative disorder; neurodegenerative disease; bacterial;
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;
XX multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
XX arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
XX antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
XX haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
XX antidiarrhetic.

OS Homo sapiens.

XX WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002WO-US005095.

XX 05-MAR-2001; 2001US-00799451.

XX (HXSQ-) HXSQ INC.

XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
XX Wehrman T, Wang J, Wang D, Dimañac RT;

XX MPI: 2002-759812/82.

XX N-PSDB; AB211891.

XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.

XX Claim 9; SEQ ID NO 1721; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a
XX nucleotide sequence selected from any of 948 sequences (AB211119-
XX AB212066) or their mature protein coding portion, active domain coding
XX protein or complementary sequences. The polynucleotides are useful for
XX identifying expressed genes or for physical mapping of human genome. The
XX encoded polypeptides (ABP68902-ABP68849) are useful as molecular weight
XX markers, as a food supplement, for generating antibodies, in medical
XX imaging, screening and diagnostic assays and for treating cell-
XX proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
XX or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
XX diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
XX platelet or coagulation disorders, wound, burns, incision, ulcers, liver
XX or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
XX arthritis, etc. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 445 AA;

Query Match 98.9%; Score 2130.5; DB 5; Length 445;
Best Local Similarity 93.9%; Pred. No. 3.2e-177;
Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

```

QY 1 MLPTALGLALSLRLALRSGAERGPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
DB 1 MLPTALGLALSLRLALRSGAERGPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
QY 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVARASQRMGDHTTGALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVARASQRMGDHTTGALVYA 120
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLEISA 180
DB 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLEISA 180
QY 181 AASAPAEKHLHFVDVDDHLIIYQELRGSLIDAMRPOQLHATEITSSGFLAMPPLLTADS 240
DB 181 AASAPAEKHLHFVDVDDHLIIYQELRGSLIDAMRPOQLHATEITSSGFLAMPPLLTADS 240
QY 241 GYVLELVPSAOPGAARQOPLGNATDWTWAGLDPTDYDALVYESNRLRLPOLILVR 300
DB 241 GYVLELVPSAOPGAARQOPLGNATDWTWAGLDPTDYDALVYESNRLRLPOLILVR 300
QY 301 TR-----PEBAGPERIVISHARPSRLVSNWAPALGSA 333
DB 301 TRPEAGPGASPGESGAPPTQALALPAPEBAGPERIVISHARPSRLVSNWAPALGSA 360

```


XX	Sequence 445 AA;	
SQL	Query Match Best Local Similarity 98.9%; Score 2130.5; DB 8; Length 445; Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1;	
QY	1 MLPTWALGLALSLRLALARSAGERGPASAPRGDLMFLDSSASVSHYEFSSRVREFVQL 60	
DB	1 MLPTWALGLALSLRLALARSAGERGPASAPRGDLMFLDSSASVSHYEFSSRVREFVQL 60	
QY	61 VAPLPLGTGALRASLVHVSRRPYTERPFQGHSSGGAADAVASAGRMGDTHGLALAYA 120	
DB	61 VAPLPLGTGALRASLVHVSRRPYTERPFQGHSSGGAADAVASAGRMGDTHGLALAYA 120	
QY	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTFVITVSTGNGFLELSA 180	
DB	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTFVITVSTGNGFLELSA 180	
QY	161 AASAPAEKHLHFVDDVDDHLIIVOELRGSLDAMRPQQLHATEITSSGFRLLAMPPLITADS 240	
DB	161 AASAPAEKHLHFVDDVDDHLIIVOELRGSLDAMRPQQLHATEITSSGFRLLAMPPLITADS 240	
QY	241 GYVVELELVSAQPGARROQLPGNATDWTWAGIDPDTVDVALVPESNVRLRPQILRV 300	
DB	241 GYVVELELVSAQPGARROQLPGNATDWTWAGIDPDTVDVALVPESNVRLRPQILRV 300	
QY	301 TR-----PEEAGPERIVISHARPSRLVSWAPALGSAA 333	
DB	301 TRPEEAGPERIVISHARPSRLVSWAPALGSAA 333	
QY	334 ALGYHVOFGPLRGGEAQRVEVPAGRNCTTLOGIAPGTAVLVTTAAFRSGRESALSAKAC 393	
DB	361 ALGYHVOFGPLRGGEAQRVEVPAGRNCTTLOGIAPGTAVLVTTAAFRSGRESALSAKAC 420	
QY	394 TPDGPRPRPRPVPRAPTGTASREP 418	
DB	421 TPDGPRPRPRPVPRAPTGTASREP 445	
RESULT 4		
AAE32501	standard; protein; 415 AA.	
ID	AAE32501	
XX	AAE32501;	
AC	24-MAR-2003 (first entry)	
DT	Mouse Willebrand Factor A domain related-protein (WARP).	
XX	Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;	
KM	extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;	
KM	gene therapy; mouse.	
XX	Mus sp.	
OS	Location/Qualifiers	
XX	1..18	
XX	/label= Signal_Peptide	
FT	Protein	
FT	19..415	
FT	/note= "Mouse mature WARP protein"	
FT	148	
FT	/note= "O-glycosylation site"	
FT	264	
FT	/note= "N-glycosylation site"	
FT	359	
FT	/note= "N-glycosylation site"	
FT	361	
FT	/note= "O-glycosylation site"	
FT	369..393	
FT	Modified-site	
FT	400	
FT	/note= "O-glycosylation site"	
XX		

PN	WO200288184-A1.	
XX		
PD	07-NOV-2002.	
XX		
PF	02-MAY-2002; 2002WO-AU000542.	
XX		
PR	02-MAY-2001; 2001AU-00004701.	
XX		
PA	(MURD-) MURDOCH CHILDRENS RES INST.	
XX		
PI	Bateman JF, Fitzgerald DJ;	
XX		
DR	WPI. 2003-111873/10.	
XX		
DR	N-PSDB; AAD50022, AAD50398.	
XX		
PT	New isolated Willebrand Factor A-Related Protein polypeptide useful for	
PT	the manufacture of a medicament in the treatment of a disease condition	
PT	of the extracellular matrix, in particular arthritis.	
XX		
PS	Claim 13; Page 74-75; 103pp; English.	
XX		
CC	The invention relates to Willebrand Factor A domain related-protein	
CC	(WARP) which is a member of von Willebrand Factor A (VA)-domain protein	
CC	superfamily of extracellular matrix (ECM) proteins. WARP is used as an	
CC	molecular marker, used for detecting a loss of ECM integrity in an animal	
CC	subject, monitoring repair, regeneration or other disease processes in an	
CC	animal subject and detecting a disease condition or a propensity for the	
CC	development of a disease condition in an animal subject. The invention is	
CC	useful for the manufacture of a medicament in the treatment of a disease	
CC	condition of the ECM. The disease condition involves the cartilage, and	
CC	is preferably arthritis. The invention is also used in gene therapy. The	
CC	present sequence is mouse WARP protein	
SQL	Sequence 415 AA;	
Query Match	76.8%; Score 1653.5; DB 6; Length 415;	
Best Local Similarity	77.5%; Pred. No. 1.4e-135;	
Matches	324; Conservative 32; Mismatches 59; Indels 3; Gaps 1;	
QY	1 MLPTWALGLALSLRLALARSAGERGPASAPRGDLMFLDSSASVSHYEFSSRVREFVQL 60	
DB	1 MLPTWALGLALSLRLALARSAGERGPASAPRGDLMFLDSSASVSHYEFSSRVREFVQL 60	
QY	61 VAPLPLGTGALRASLVHVSRRPYTERPFQGHSSGGAADAVASAGRMGDTHGLALAYA 120	
DB	61 VAPLPLGTGALRASLVHVSRRPYTERPFQGHSSGGAADAVASAGRMGDTHGLALAYA 120	
QY	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTFVITVSTGNGFLELSA 180	
DB	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTFVITVSTGNGFLELSA 180	
QY	161 AASAPAEKHLHFVDDVDDHLIIVOELRGSLDAMRPQQLHATEITSSGFRLLAMPPLITADS 240	
DB	161 AASAPAEKHLHFVDDVDDHLIIVOELRGSLDAMRPQQLHATEITSSGFRLLAMPPLITADS 240	
QY	241 GYVVELELVSAQPGARROQLPGNATDWTWAGIDPDTVDVALVPESNVRLRPQILRV 300	
DB	241 GYVVELELVSAQPGARROQLPGNATDWTWAGIDPDTVDVALVPESNVRLRPQILRV 300	
QY	301 TRPEEAGPERIVISHARPSRLVSWAPALGSAAALGYHVOFGPLRGGEAQRVEVPAGRN 360	
DB	301 TRPEEAGPERIVISHARPSRLVSWAPALGSAAALGYHVOFGPLRGGEAQRVEVPAGRN 360	
QY	361 TTVQGLTCTTYLVTTAAFRSGRESALSAKACTPDGPRPRPVPRAPTGTASREP 418	
DB	361 TTVQGLTCTTYLVTTAAFRSGRESALSAKACTPDGPRPRPVPRAPTGTASREP 415	
RESULT 5		
AAW86326	standard; protein; 421 AA.	
ID	AAW86326	
XX	AAW86326;	
AC		

XX 01-MAR-1999 (first entry)
XX Kidney injury associated molecule HM059 protein.
DE Kidney injury associated molecule; kidney injury related molecule; KIM;
XX Kidney injury associated molecule; kidney injury related molecule; KIM;
XX tissue growth promotion; regeneration; renal condition;
XX acute renal failure; acute nephritis; tumour.
OS Ratus sp.
XX WO9853071-A1.
XX 26-NOV-1998.
XX 22-MAY-1998; 98WO-US010547.
XX 23-MAY-1997; 97US-0047490P.
XX 23-MAY-1997; 97US-0047491P.
XX (BIO) BIOGEN INC.
XX Sanicola-Nadel M, Hession CA, Wei H, Cate RL;
XX WPI; 1999-045312/04.
XX N-PSDB; AAV80611.
XX Kidney injury-associated molecule, KIM, polypeptides - upregulated in
XX injured or regenerating tissues, useful to promote tissue growth and
XX regeneration, especially to treat renal conditions.
XX Claim 17; Page 135-137; 213pp; English.
XX The present sequence represents a kidney injury associated molecule (KIM)
XX protein. KIM proteins can be administered therapeutically by expressing
XX KIM encoding polynucleotides, to promote growth and/or survival of
XX damaged tissue (e.g. renal tissue), since the KIM proteins are
XX upregulated in injured or regenerating (especially renal) tissues. KIM
XX fusion proteins, conjugates, antibodies and vectors can also be used
XX therapeutically, e.g. these or the KIM proteins may be included with an
XX acceptable carrier in pharmaceutical compositions, useful for therapy/
XX prophylaxis of conditions associated with dysfunction/dysregulation of
XX KIM genes or proteins, especially renal diseases or impairments of renal
XX function in humans (e.g. acute renal failure, acute nephritis). The
XX polynucleotides can be used to produce antisense sequences which, when
XX internalised into cells, can disrupt expression of a cellular KIM gene,
XX also useful in therapy (e.g. to block the growth of tumours dependent on
XX KIM for growth) or compositions. The proteins and polynucleotides are
XX useful diagnostically e.g. to detect and quantify renal injury/disease
XX (indicative of increased risk, or presence of, renal injury or impaired
XX function), or abnormal responses to tissue injury (indicative of
XX increased risk, or presence of, an autoimmune response or abnormal tissue
XX growth arising from/affecting renal tissue). The proteins can also be
XX used to locate KIM-producing cells (especially specific loci, e.g. tissue
XX masses abnormally producing/expressing KIM such as tumours arising
XX from/affecting renal tissue), by contacting cells with an imageable KIM-
XX binding reagent and imaging reagent accumulation
XX
XX Sequence 421 AA;
SQ

Query Match 76.3%; Score 1643.5; DB 2; Length 421;
Best Local Similarity 77.8%; Pred. No. 1,1e-134;
Matches 325; Conservative 31; Mismatches 59; Indels 3; Gaps 1;

QY 1 MLEPWLGLALSLRLALARSAGARPAPAPRODLMFLDLSASVSHYEFRRVFGQT 60
DB 7 MLEPWLGLALSLRLALARSAGARPAPAPRODLMFLDLSASVSHYEFRRVFGQT 66
QY 61 VAPLPGLTALSLRLALARSAGARPAPAPRODLMFLDLSASVSHYEFRRVFGQT 120
DB 67 VAPLPGLTALSLRLALARSAGARPAPAPRODLMFLDLSASVSHYEFRRVFGQT 126
QY 121 KEQLFAEASGARPGVPRVLTWVTDGSSDVGPPMQLKDLGVTVEIVSTGRGNFLELSA 180

DB 127 KEQLFAEASGARPGVPRVLTWVTDGSSDVGPPMQLKDLGVTVEIVSTGRGNFLELSA 186
QY 181 AASAPAEKHLHFDVDDHIIIVQELRGSITLDMRPOQLHATEITSSGFLAPPLTTADS 240
DB 187 AASAPAEKHLHFDVDDHIIIVQELRGSITLDMRPOQLHATEITSSGFLAPPLTTADS 246
QY 241 GYVLELVPSAQPGARQQLPGNATDWMAGLDPTDVALVESNRLRLPQLLRVR 300
DB 247 GYVLELVPSAQPGARQQLPGNATDWMAGLDPTDVALVESNRLRLPQLLRVR 306
QY 301 TPPEAGPERIVISHPARPSRLRVSWAPALGSAALGIVQFPPLGGEQRVEVPAGRC 360
DB 307 TLQEEAGPERIVISHPARPSRLRVSWAPALGSAALGIVQFPPLGGEQRVEVPAGRC 366
QY 361 TLQEEAGPERIVISHPARPSRLRVSWAPALGSAALGIVQFPPLGGEQRVEVPAGRC 418
DB 367 TLQEEAGPERIVISHPARPSRLRVSWAPALGSAALGIVQFPPLGGEQRVEVPAGRC 421

RESULT 6
AAB42581
ID AAB42581 standard; protein; 299 AA.
XX AAB42581;
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF2345 polypeptide sequence SEQ ID NO:4690.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
XX anticonvulsant; osteoporotic; antiarthritic; immunosuppressant; cardiac;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX thrombosis; cartilage damage; antiinflammatory disease; coagulation;
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US008621.
XX 31-MAR-1999; 99US-0127607P.
XX 02-APR-1999; 99US-0127636P.
XX 05-APR-1999; 99US-0127728P.
XX 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AACT6790.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX Claim 11; Page 3880-3881; 5507pp; English.
XX AACT7446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;
CC cardiact; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antineumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 299 AA;

SQ Query Match 63.2%; Score 1362; DB 3; Length 299;

Best Local Similarity 93.2%; Pred. No. 2.7e-110; Mismatches 16; Indels 4; Gaps 2;

Matches 274; Conservative 0; Mismatches 16; Indels 4; Gaps 2;

QY 3 PWTALGLALSLRLALRSAGERPPASAPRGDLMFLDSSASVSHYEFGRVREFVQQLVA 62
DB 10 PLDGLARLALSLRW---RWRTPDCCPPASABE-DLMFLDSSASVSHYEFGRVREFVQQLVA 65

QY 63 PLPLGTGALRASLVHVGSRPYTEFFPGQSSGEEAODAVRASQRMGDHTHTGLALVYAAE 122
DB 66 PLPLAPXALRASLVHVGSRPYTEFFPGQSSGEEAODAVRASQRMGDHTHTGLALVYAKE 125

QY 123 QLFPAASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDIQVTVFIVSTGRGNFLLELSAA 182
DB 126 QLFPAASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDIQVTVFIVSTGRGNFLLELSAA 185

QY 183 SAPAEKHLHFVDDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 242
DB 186 SAPAEKHLHFVDDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 245

QY 243 VYLELVPSAQPGARPOQLPGNATDWMAGLDPTDIDYDALVPSVNRLLRPQI 296
DB 246 VYLELVPSAQPGARPOQLPGNATDWMAGLDPTDIDYDALVPSVNRLLRPQI 299

RESULT 7
AAB88340

ID AAB88340 standard; protein; 285 AA.

XX AAB88340;

DT 23-MAY-2001 (first entry)

DE Human membrane or secretory protein clone PSEC0053.

XX Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes.

OS Homo sapiens.

XX Homo sapiens.

PN EP1067182-A2.

PD 10-JAN-2001.

PF 07-JUL-2000; 2000EP-00114090.

PR 08-JUL-1999; 99JP-00194179.

PR 11-JAN-2000; 2000JP-00118775.

PR 02-MAY-2000; 2000JP-00183765.

XX (HELI-) HELIX RES INST.

PI Ota T, Ieogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX WPI; 2001-093989/11.
DR N-PSDB; AAF93767.

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development.

PS Claim 1; SEQ ID NO 48; 609pp + Sequence Listing; English.

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88419, included in the invention are primers AAF93917 - AAF94295 and
CC AAF92232 - AAF62235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and antagonists
CC may also be used as therapeutic agents to down-regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme
CC linked immunosorbent assay (ELISA). Examples of diseases which may be
CC treated include rheumatoid arthritis and diabetes

SQ Sequence 285 AA;

Query Match 59.8%; Score 1288.5; DB 4; Length 285;

Best Local Similarity 96.6%; Pred. No. 6.7e-104; Mismatches 2; Indels 5; Gaps 1;

Matches 256; Conservative 2; Mismatches 2; Indels 5; Gaps 1;

QY 1 MLPTALGLALSLRLALRSAGERPPASAPRGDLMFLDSSASVSHYEFGRVREFVQQL 60
DB 1 MLPTALGLALSLRLALRSAGERPPASAPRGDLMFLDSSASVSHYEFGRVREFVQQL 60

QY 61 VAPPLGTGALRASLVHVGSRPYTEFFPGQSSGEEAODAVRASQRMGDHTHTGLALVYAA 120
DB 61 VAPPLGTGALRASLVHVGSRPYTEFFPGQSSGEEAODAVRASQRMGDHTHTGLALVYAA 120

QY 121 KEQLFAASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDIQVTVFIVSTGRGNFLLELSA 180
DB 121 KEQLFAASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDIQVTVFIVSTGRGNFLLELSA 180

QY 181 AASAPAEKHLHFVDDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 240
DB 181 AASAPAEKHLHFVDDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 240

QY 241 GYVVLVPSAQPG-----AARQ 260
DB 241 GYVVLVPSAQPGCKTPAARER 265

RESULT 8
ADY63045

ID ADY63045 standard; protein; 285 AA.

XX ADY63045;

DT 02-JUN-2005 (first entry)

DE Human clone PSEC0053 protein, SEQ ID 48.

XX Gene therapy.

OS Homo sapiens.
XX
XX EP1514933-A1.
XX
XX 16-MAR-2005.
XX
XX 07-JUL-2000; 2004EP-00027228.
XX
XX 08-JUL-1999; 99JP-00194179.
XX 11-JAN-2000; 2000JP-00118775.
XX 02-MAY-2000; 2000JP-00183766.
XX 07-JUL-2000; 2000EP-00114090.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
XX WPI; 2005-203865/22.
XX
XX N-PSDB; ADY63044.
XX
XX Novel isolated polynucleotide encoding human secretory proteins or
XX PT membrane proteins, useful for examination and diagnosis of abnormality of
XX PT human secretory proteins.
XX
XX PS Disclosure; SEQ ID NO 48; 1240bp; English.
XX
XX The present invention relates to novel human secretory proteins or
XX CC membrane proteins, and their coding sequences. The present sequence is
XX CC one such protein sequence. The coding sequences of the invention are
XX CC useful for examination and diagnosis of abnormality of the human
XX CC secretory proteins and in gene therapy methods. The coding sequences and
XX CC proteins are useful as candidates for medicines or as target molecules
XX CC for developing medicines. Antibodies against the proteins of the
XX CC invention are useful for treating diseases that are associated with the
XX CC proteins. Note: The sequence data for this patent did not form part of
XX CC the printed specification, but was obtained from sequence information
XX CC supplied by the European Patent Office.
XX
XX SQ Sequence 285 AA;
XX
XX Query Match 59.8%; Score 1288.5; DB 9; Length 285;
XX Best Local Similarity 96.6%; Pred. No. 6.7e-104;
XX Matches 256; Conservative 2; Mismatches 2; Indels 5; Gaps 1;
XX
XX QY 1 MLFWTALGLALSLRLALARSAGRGPPASAPRGDLFLDSSASVSHYEFRRVFEVQGL 60
XX 1 MLFWTALGLALSLRLALARSAGRGPPASAPRGDLFLDSSASVSHYEFRRVFEVQGL 60
XX DB 1 MLFWTALGLALSLRLALARSAGRGPPASAPRGDLFLDSSASVSHYEFRRVFEVQGL 60
XX
XX QY 61 VAPLPIGTGALRASLVHVGSRPYTFEPFGQSSGEAAQDAVRASAOQRMGDTHTGLALVYA 120
XX 61 VAPLPIGTGALRASLVHVGSRPYTFEPFGQSSGEAAQDAVRASAOQRMGDTHTGLALVYA 120
XX DB 61 VAPLPIGTGALRASLVHVGSRPYTFEPFGQSSGEAAQDAVRASAOQRMGDTHTGLALVYA 120
XX
XX QY 121 KEQLEFASAGARGVKKVLMWTDGSSDPVGPMPQELDGLGVTVIVSGRNFLELSA 180
XX 121 KEQLEFASAGARGVKKVLMWTDGSSDPVGPMPQELDGLGVTVIVSGRNFLELSA 180
XX DB 121 KEQLEFASAGARGVKKVLMWTDGSSDPVGPMPQELDGLGVTVIVSGRNFLELSA 180
XX
XX QY 181 AASAPAEKHLHFDVVDLHITVOELRGSILNARPOOLNATETITSSGFRMLPPLLTADS 240
XX 181 AASAPAEKHLHFDVVDLHITVOELRGSILNARPOOLNATETITSSGFRMLPPLLTADS 240
XX DB 181 AASAPAEKHLHFDVVDLHITVOELRGSILNARPOOLNATETITSSGFRMLPPLLTADS 240
XX
XX QY 241 GYVVLGLVPSAQOGGCTPPAAAMER 265
XX 241 GYVVLGLVPSAQOGGCTPPAAAMER 265
XX DB 241 GYVVLGLVPSAQOGGCTPPAAAMER 265
XX
XX RESULT 9
XX AAB87344 ID AAB87344 standard; protein; 215 AA.
XX
XX AC AAB87344;
XX
XX DT 22-MAY-2001 (first entry)
XX

DE Human gene 3 encoded secreted protein HNTB078, SEQ ID NO:85.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
XX KW foetal abnormality; developmental abnormality; haematopoietic disorder;
XX KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX KW inflammation; allergy; neurological disorder; Alzheimer's disease;
XX KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
XX KW cardiovascular disorder; angiotensin disorder; kidney disorder;
XX KW gastrointestinal disorder; pregnancy-related disorder;
XX KW endocrine disorder; infection; wound healing; vulnerability; cell culture;
XX KW chemotaxis; food additive; binding partner identification.
XX
XX OS Homo sapiens.
XX
XX PN WO200118022-A1.
XX
XX PD 15-MAR-2001.
XX
XX PF 31-AUG-2000; 2000WO-US024008.
XX
XX PR 03-SEP-1999; 99US-0152315P.
XX PR 03-SEP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Ni J, Baker KP, Birse CE, Piscella M, Komatsu GA, Rosen CA;
XX PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
XX PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
XX DR WPI; 2001-203081/20.
XX
XX N-PSDB; AAF91860.
XX
XX PT Nucleic acid molecules encoding human secreted proteins, used in
XX PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
XX PT Parkinson's diseases and cancers.
XX
XX PS Claim 11; Page 532-533; 607bp; English.
XX
XX AA AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
XX CC proteins, and AAB87342-AAB87413 represent the proteins they encode.
XX CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
XX CC and their corresponding secreted proteins are useful for preventing,
XX CC treating or ameliorating medical conditions, e.g., by protein or gene
XX CC therapy. Pathological conditions can be diagnosed by determining the
XX CC amount of the new protein in a sample or by determining the presence of
XX CC mutations in the new genes. Specific uses are described for each of the
XX CC 52 genes, based on the tissues in which they are most highly expressed,
XX CC and include developing products for the diagnosis or treatment of
XX CC proliferative disorders, cancer, tumours, foetal and developmental
XX CC abnormalities, haematopoietic disorders, diseases of the immune system,
XX CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX CC allergies, neurological disorders (e.g., Alzheimer's disease,
XX CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
XX CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX CC cardiovascular disorders, angiotensin disorders, kidney disorders,
XX CC gastrointestinal disorders, pregnancy-related disorders, endocrine
XX CC disorders, and infections. The proteins can also be used to aid wound
XX CC healing and epithelial cell proliferation, to prevent skin aging due to
XX CC sunburn, to maintain organs before transplantation, for supporting cell
XX CC culture of primary tissues, to regenerate tissues, to identify their
XX CC cognate ligands or binding partners, and in chemotaxis, and can be used
XX CC as a food additive or preservative to modify storage properties.
XX CC Antibodies specific for a protein of the invention can be used in
XX CC alleviating symptoms associated with the disorders mentioned above, and
XX CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX CC immunosorbent assay (ELISA). The present sequence represents a human
XX CC secreted protein of the invention
XX
XX SQ Sequence 215 AA;
XX
XX Query Match 50.3%; Score 1084; DB 4; Length 215;
XX Best Local Similarity 99.5%; Pred. No. 3.4e-86;
XX

Matches	214;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0
QY		1	MLPFWALGIALSLRLALRSGAERGPPASAPGDIIMFLIDSSASVSHYFSVRREVGQL	60					
Db		1	MLPFWAXGIALSLRLALRSGAERGPPAPRGDLMFLIDSSASVSHYFSVRREVGQL	60					
QY		61	VAPFLVGTGALPASLVHVSRRPYTEPPFQGHSSGGAQDAVPAQAQRMGDTHTGLALVYA	120					
Db		61	VAPFLVGTGALPASLVHVSRRPYTEPPFQGHSSGGAQDAVPAQAQRMGDTHTGLALVYA	120					
QY		121	KEQLFAEASGARPGVPKVLVWTTDGSSPVCPMPOMELDLGVTYVIVSTGKGNFLEISA	180					
Db		121	KEQLFAEASGARPGVPKVLVWTTDGSSPVCPMPOMELDLGVTYVIVSTGKGNFLEISA	180					
QY		181	AASAPAEKHLHFVDVDDLHIIVQELRGSLIDAMP	215					
Db		181	AASAPAEKHLHFVDVDDLHIIVQELRGSLIDAMP	215					
RESULT 10									
ABG65347		ID	ABG65347	standard; protein; 215 AA.					
AC			ABG65347;						
DT			27-AUG-2002 (first entry)						
DE			Human albumin fusion protein #2022.						
KM			Albumin fusion protein; therapeutic protein X; human albumin; HA;						
KM			human serum albumin; HSA; cancer; reproductive disorder;						
KM			digestive disorder; immune disorder; endocrine disorder;						
KM			haematopoietic disorder; neural disorder; connective disorder;						
KM			cytosolic; antifertility; antiinflammatory; anticancer;						
KM			immunomodulator; anti-HIV; antidiabetic; haemostatic; nocrotropic;						
KM			neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;						
XX			osteopathic; antiarthritic.						
OS			Homo sapiens.						
OS			Synthetic.						
PN			WO200177137-A1.						
PD			18-OCT-2001.						
PF			12-APR-2001; 2001WO-US011988.						
PR			12-APR-2000; 2000US-0229358P.						
PR			25-APR-2000; 2000US-0199384P.						
PR			21-DEC-2000; 2000US-0256931P.						
PA			(HUMA-) HUMAN GENOME SCI INC.						
PI			Rosen CA, Haseeltine WA;						
DR			WPI; 2002-010886/01.						
PT			New fusion protein for treating disease e.g. diabetes comprises an						
PT			albumin fused to a therapeutic protein.						
PS			Claim 1; Page 1935; 2102pp; English.						
CC			The present invention relates to albumin fusion proteins comprising a						
CC			therapeutic protein X and human albumin (HA, also known as human serum						
CC			albumin, HSA). The proteins are useful for treating a disease or disorder						
CC			that may be modulated by therapeutic protein X. The albumin extends the						
CC			shelf-life of protein X, and may increase its biological in vitro/in vivo						
CC			activity. The protein X, and may increase its biological in vitro/in vivo						
CC			such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's						
CC			disease, ulcerative colitis), immune disorders (e.g. acquired						
CC			immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),						
CC			haematopoietic disorders, neural disorders (e.g. Alzheimer's,						
CC			Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,						

Query Match	50.3%	Score 1084	DB 5	Length 215
Best Local Similarity	99.5%	Pred. No. 3,4e-86		
Matches 214	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Qy	1	MLPWTALGLALSTRLALASGAERPPASAPRGDMLFLDSSASVSHYFSGRYAEFPYGOL	60	
Db	1	MLPWTAXGLALSTRLALASGAERPPASAPRGDMLFLDSSASVSHYFSGRYAEFPYGOL	60	
Qy	61	VAPLPFGTGLRASLVHVGSRPYTEFFPGQHSGBAAQDAVRASAOQMGDTHTGLALVYA	120	
Db	61	VAPLPFGTGLRASLVHVGSRPYTEFFPGQHSGBAAQDAVRASAOQMGDTHTGLALVYA	120	
Qy	121	KEQLFAASGARPGVQKVLWVTTDGGSDPVGPPQELKDIQVTVFVSTGRGFLELSA	180	
Db	121	KEQLFAASGARPGVQKVLWVTTDGGSDPVGPPQELKDIQVTVFVSTGRGFLELSA	180	
Qy	181	AASAPAEKHLHFDVDDHLITVOELRGSLDAMP	215	
Db	181	AASAPAEKHLHFDVDDHLITVOELRGSLDAMP	215	
RESULT 11				
ADL78614				
ID	ADL78614	standard; protein; 215 AA.		
AC	ADL78614			
XX				
DT	20-MAY-2004	(first entry)		
DE		Albumin fusion protein related therapeutic protein X, SEQ ID NO 2096.		
XX				
KM		albumin fusion protein; cytostatic; antinaemic; antiarthritic;		
KM		antiaesthetic; anti-HIV; immunosuppressive; antiinflammatory;		
KM		antibacterial; antibacterial; osteopathic; dermatological; angiot;		
KM		immunomodulator; antiarrhythmic; cardiac; nootropic; antilipemic;		
KM		nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer		
KM		antidiabetic; anabolic; hypertensive; vulnerary; gene therapy; cancer;		
KM		reproductive system disorder; therapeutic protein.		
XX				
OS		Unidentified.		
XX				
PN	US2004010134-A1.			
XX				
PD	15-JAN-2004.			
XX				
PF	12-APR-2001; 2001US-00833245.			
XX				
PR	12-APR-2000; 2000US-0229358P.			
PR	25-APR-2000; 2000US-0199384P.			
PR	21-DEC-2000; 2000US-0256931P.			
XX				
PA	(ROSE/) ROSEN C A.			
PA	(HASE/) HASELTINE W A.			
XX				
PI	Rosen CA, Haseltine WA;			
XX				
XX	WPI; 2004-090519/09.			
XX				
XX				
PT	New albumin fusion proteins; useful for diagnosing, treating, preventing			
PT	or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,			
PT	asthma, inflammatory bowel disease or Alzheimer's disease.			
XX				
PS	Disclosure; SEQ ID NO 2096; 279pp; English.			
CC	The invention relates to a novel albumin fusion protein. The invention			
CC	further relates to: a composition comprising the albumin fusion protein			
CC	and a pharmaceutical carrier; a kit comprising the composition of the			
CC	albumin fusion protein formula; a method of treating a disease or			

disorder in a patient comprising the step of administering the albumin fusion protein; a method of treating a patient with a disease or disorder that is modulated by Therapeutic Protein X, or its fragment or variant; a method of extending the shelf life of Therapeutic protein X, or its fragment or variant; a nucleic acid molecule comprising a polynucleotide sequence encoding the albumin fusion protein; a vector comprising the nucleic acid molecule of the albumin fusion protein; and a host cell comprising the nucleic acid molecule of the albumin fusion protein. The albumin fusion protein and its compositions have the following activities: cytostatic, antineoplastic, antiarthritic, antitachytic, anti-HIV, immunosuppressive, antiinflammatory, antiproliferative, antibacterial, osteoprotective, dermatological, antigout, immunomodulatory, antiarhythmic, cardiatic, neurotropic, antilipemic, nephroprotective, uropathic, neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive, and vulnerary. The albumin fusion protein nucleic acid may be used in gene therapy to treat disorders. The albumin fusion protein is useful for diagnosing, treating, preventing or ameliorating diseases or disorders comprising indication: X. The diseases or disorders include: cancer (e.g. leukaemia, colon, bone, breast, liver or lung cancer), immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute lymphocytic anaemia, multiple myeloma, arthritis, psoriasis or Lyme autoimmune disease, inflammatory bowel disease), psoriasis or Lyme disease), reproductive system disorders (e.g. prostaticitis, inguinal hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-Leydig tumour), musculoskeletal diseases (e.g. giant cell tumours, Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy or cacthexia), cardiovascular disease (e.g. rhabdomyomas, heart disease, arrhythmia, cardiac arrest, heart valve disease, hypernatremia or hyponatremia), mixed foetal diseases (e.g. foetal alcohol syndrome, Down's syndrome, Patau syndrome, Turner's syndrome, Apert syndrome or Tay-Sachs disease), excretory diseases (e.g. urinary incontinence, urinary tract infections or renal disorders), neural or sensory disease (e.g. Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis, cerebellar ataxia, attention deficit disorder, autism or obsessive compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or occupational lung disease), endocrine diseases (e.g. diabetes, Addison's disease or glomerulonephritis), digestive diseases (e.g. portal hypertension, irritable bowel disease, gastric atrophy or pancreatitis) or connective tissue or epithelial diseases (e.g. Crohn's disease, scleroderma, wound healing or epidermolysis bullosa). This sequence represents a therapeutic protein X relating to the albumin fusion protein of the invention. The sequence listing data for this specification was downloaded from the USPTO website.

Sequence 215 AA;

Query Match	50.3%	Score 1084	DB 8	Length 215
Best Local Similarity	99.5%	Pred. No. 3.4e-86		
Matches 214; Conservative	0	Mismatches 1	Indels 0	Gaps 0

Db 1 MLPWYXGLATLRLALNRSAGERGPAPAFKQDLMFLDSSASVSHYERSRREYVQL 60

QY 61 VAPLPLGTGATLPAALVHVGSRPYTEFPFGQSHSSGEAAODAVRASQRMGDTHITGLVYA 120

Db 61 VAPLPLGTGATLPAALVHVGSRPYTEFPFGQSHSSGEAAODAVRASQRMGDTHITGLVYA 120

QY 121 KEQLFAPASGARPGVPKVLWWTDDGSSSPVCPGPMELDIDGTVTIVSTGRGNFLIELSA 180

Db 121 KEQLFAPASGARPGVPKVLWWTDDGSSSPVCPGPMELDIDGTVTIVSTGRGNFLIELSA 180

QY 181 AASAPAEKRLHFFVDVDDLHIIVOELNGSILLDAMP 215

Db 181 AASAPAEKRLHFFVDVDDLHIIVOELNGSILLDAMP 215

RESULT	12
AA	AA87418
ID	AA87418 standard; protein; 242 AA
XX	
AC	AA87418;
XX	

DT 22-MAY-2001 (first entry)
XX
XX Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
DE
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; achma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery; cell culture;
KW chemocaxis; food additive; binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200118022-A1.
XX
PD 15-MAR-2001.
XX
PF 31-AUG-2000; 2000WO-US024008.
XX
PR 03-SEP-1999; 99US-0152315P.
XX 03-SEP-1999; 99US-0152317P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Piscella M, Komatoulis GA, Rosen CA,
PI Sopet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW,
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
DR WPI; 2001-203081/20.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
PT
XX
PS Disclosure; Page 18; 607pp; English.
XX
XX AAEP91858-AAEP1929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, achma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemocaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX Sequence 242 AA;

Query Match 50.3%; Score 1084; DB 4; Length 242;

Best Local Similarity 99.5%; Pred. No. 4e-86;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLPWTALGLAISLRALARSAGRGPPASAPRGDMLFLDSSASVSHYEFSSVRFVQOL 60
DB 28 MLPWTALGLAISLRALARSAGRGPPASAPRGDMLFLDSSASVSHYEFSSVRFVQOL 87
QY 61 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHTGALVYA 120
DB 88 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHTGALVYA 147
QY 121 KEOLFPAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180
DB 148 KEOLFPAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 207
QY 181 AASAPAEKRLHFPVDVDDHLITVOELRGSILDMARP 215
DB 208 AASAPAEKRLHFPVDVDDHLITVOELRGSILDMARP 242

RESULT 13
AAB87424
ID AAB87424 standard; protein; 226 AA.
XX
AC AAB87424;
XX
DT 22-MAY-2001 (first entry)
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery; cell culture;
KW chemotaxis; food additive; binding partner identification.

OS Homo sapiens.
XX
PN WO200118022-A1.
XX
PD 15-MAR-2001.
XX
PE 31-AUG-2000; 2000MO-US024008.
XX
PR 03-SEP-1999; 99US-0152315P.
PR 03-SEP-1999; 99US-0152317P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsu G, Rosen CA;
PI Sopet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
DR WPI, 2001-203081/20.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
PS Disclosure; Page 18; 607pp; English.
XX
CC AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of

CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorder, angiogenic disorders, kidney disorder,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX

SO Sequence 226 AA;

Query Match 50.0%; Score 1077; DB 4; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.5e-85;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPWTALGLAISLRALARSAGRGPPASAPRGDMLFLDSSASVSHYEFSSVRFVQOL 60
DB 14 MLPWTALGLAISLRALARSAGRGPPASAPRGDMLFLDSSASVSHYEFSSVRFVQOL 73
QY 61 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHTGALVYA 120
DB 74 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHTGALVYA 133
QY 121 KEOLFPAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180
DB 134 KEOLFPAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 193
QY 181 AASAPAEKRLHFPVDVDDHLITVOELRGSILDMARP 213
DB 194 AASAPAEKRLHFPVDVDDHLITVOELRGSILDMARP 226

RESULT 14
AAB03654
ID AAB03654 standard; protein; 185 AA.
XX
AC AAB03654;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human extracellular matrix and cell adhesion molecule-18 (XNAD-18).
XX
KW Human; extracellular matrix and cell adhesion molecule; XNAD;
KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
KW infection; cell proliferative disorder; actinic keratosis; myeloma;
KW arteriosclerosis; nootropic; anticonvulsant; antithyroid; nephrotoxic;
KW neuroprotective; dermatological.

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1..171

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 13:10:29 ; Search time 85.2776 Seconds
(without alignments)
2048.049 Million cell updates/sec

Title: US-10-699-035A-6
Perfect score: 2154
Sequence: 1 MIPWTLGLALSRLALARS.....RPRPRVPRAPPTGTASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubpaa/us07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/us08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubpaa/us09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubpaa/us10_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/us10_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubpaa/us11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2154	100.0	418	4	US-10-699-035A-6
2	2154	100.0	418	4	US-10-699-035A-20
3	1653.5	76.8	415	4	US-10-699-035A-4
4	1653.5	76.8	415	4	US-10-699-035A-21
5	1084	50.3	215	3	US-09-789-561-85
6	1084	50.3	215	3	US-09-833-245-2096
7	1084	50.3	215	5	US-10-883-936-85
8	1084	50.3	242	3	US-09-789-561-159
9	1084	50.3	242	5	US-10-883-936-159
10	1077	50.0	226	3	US-09-789-561-165
11	1077	50.0	226	5	US-10-883-936-165
12	913	42.6	185	4	US-10-149-819-18
13	913	42.6	180	4	US-10-699-035A-2
14	772	35.7	181	4	US-10-699-035A-31
15	768	35.8	180	4	US-10-699-035A-8
16	424	19.7	3063	5	US-10-177-293-61
17	424	19.7	3063	5	US-10-631-467-918
18	421.5	19.6	3063	3	US-09-918-715-257
19	421.5	19.6	3063	4	US-10-177-293-61
20	421.5	19.6	3063	4	US-10-301-822-26
21	421.5	19.6	3063	4	US-10-474-794-257
22	421.5	19.6	3063	5	US-10-979-159-257
23	421.5	19.6	3118	5	US-10-220-335-287
24	407	18.9	3067	5	US-10-631-467-1618
25	392.5	18.2	1207	4	US-10-408-765A-1591
26	392.5	18.2	1297	4	US-10-187-975-102
27	382.5	17.8	2502	5	US-10-741-600-1627

28	382.5	17.8	2944	5	US-10-723-860-1785	Sequence 1785, Ap
29	382.5	17.8	2944	5	US-10-741-600-1629	Sequence 1629, Ap
30	366	17.0	87	4	US-10-699-035A-40	Sequence 40, Ap1
31	365	16.9	87	4	US-10-699-035A-36	Sequence 36, Ap1
32	360	16.7	1899	5	US-10-631-467-919	Sequence 919, Ap1
33	326	15.1	1780	4	US-10-115-479-10	Sequence 10, Ap1
34	272.5	12.7	176	3	US-09-976-782-38	Sequence 38, Ap1
35	268.5	12.5	843	4	US-10-408-765A-17	Sequence 17, Ap1
36	265.5	12.3	182	4	US-10-699-035A-22	Sequence 22, Ap1
37	264	12.3	755	3	US-09-919-497-57	Sequence 57, Ap1
38	263.5	12.2	184	4	US-10-699-035A-27	Sequence 27, Ap1
39	259	12.0	580	5	US-10-723-860-2769	Sequence 2769, Ap
40	253.5	11.8	776	4	US-10-000-512-8	Sequence 8, Ap1
41	253.5	11.8	776	4	US-10-074-566-8	Sequence 8, Ap1
42	253.5	11.8	782	4	US-10-428-275-176	Sequence 176, Ap1
43	253.5	11.8	794	3	US-09-833-245-1292	Sequence 1292, Ap
44	253.5	11.8	914	4	US-10-428-275-156	Sequence 156, Ap
45	253.5	11.8	915	3	US-09-909-320-34	Sequence 34, Ap1

ALIGNMENTS

RESULT 1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
US-10-699-035A-6																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; Sequence 6, Application US/10699035A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; Publication No. US20040214349A1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
GENERAL INFORMATION:																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; APPLICANT: Bateman, John																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; APPLICANT: Fitzgerald, David																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; TITLE OF INVENTION: A Molecular Marker																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; FILE REFERENCE: A36056 PCT USA A 071838.0142																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; CURRENT APPLICATION NUMBER: US/10/699,035A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; CURRENT FILING DATE: 2003-10-31																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; PRIOR APPLICATION NUMBER: PCT/AU02/00542																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; PRIOR FILING DATE: 2002-05-02																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; PRIOR APPLICATION NUMBER: AU PR4701/01																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; PRIOR FILING DATE: 2001-05-02																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; NUMBER OF SEQ ID NOS: 40																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; SOFTWARE: PaateSeq for Windows Version 4.0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; SEQ ID NO 6																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; LENGTH: 418																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; TYPE: PRT																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
; ORGANISM: Homo sapiens																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
US-10-699-035A-6																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Query Match																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Best Local Similarity 100.0%; Score 2154; DB 4; Length 418;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Best Local Similarity 100.0%; Pred. No. 2e-174;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
QY	1	MIPWTALGLALSRLALARS	GAE	R	G	P	P	A	S	A	S	V	S	H	E	F	S	R	V	E	F	V	G	L																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						</


```
Db      1 MFWTAFSVALSLRLALARSISIERGSTASDPQDILLFLDSSASVSHYFSRVEFGQL 60
Qy      61 VAPLPGTGLALASLVHVSRRPYTEPPFGQHSSEGAQDAVNASARMDGTHGLALVYA 120
Db      61 VMTMFGPALRLASLVHVSQPTTEFTFDYSSGQALRLAKVAPRMDGTHGLALVYA 120
Qy      121 KEOLFPAASGARPGVVKVLMWTDGSSDPVGPMPQELKDGLGVTVIVSTGRGNFLELSA 180
Db      121 KEOLFPAASGARPGVVKVLMWTDGSSDPVGPMPQELKDGLGVTVIVSTGRGNFLELSA 180
Qy      181 AASAPAEKHLHFVDVDDLHIIVQELRGSLTDAMQPQOLHASELSSGFRLSMPLLTADS 240
Db      241 GYYVLELVSAOAGARROOLPGNATDWMAGIDPPTDVALVPSNRLLRPOLLYR 300
Qy      241 GYYVLELVSAOAGARROOLPGNATDWMAGIDPPTDVALVPSNRLLRPOLLYR 300
Db      301 TRPEAGPERIVISHARPSRLRVSMAPALGSAALGVHVFGLRGSEAGQREVPAGRNC 360
Qy      301 TRPEAGPERIVISHARPSRLRVSMAPALGSAALGVHVFGLRGSEAGQREVPAGRNC 360
Db      301 TIQEEAGPERIVISHARPSRLRVSMAPALGSAALGVHVFGLRGSEAGQREVPAGRNC 360
Qy      361 TTLQGLAPGTAYLVYTTAAFRSGESALSAKACTPPGPPRPVRAPIPTGASREP 418
Db      361 TTVOGLTPTCTTYLVYTTAAFRSGORALSAKACTASGARTR--APQSMKPEAGPREP 415

RESULT 5
US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Nt et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-85

Query Match      50.3%; Score 1084; DB 3; Length 215;
Best Local Similarity 99.5%; Pred. No. 8.4e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Db      181 AASAPAEKHLHFVDVDDLHIIVQELRGSLTDAMRP 215

RESULT 6
US-09-833-245-2096
; Sequence 2096, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2096
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

Query Match      50.3%; Score 1084; DB 3; Length 215;
Best Local Similarity 99.5%; Pred. No. 8.4e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

SEQ ID NO 85
LENGTH: 215
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (7)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-85

Query Match 50.3%; Score 1084; DB 5; Length 215;
Best Local Similarity 99.5%; Pred. No. 8.4e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLPWTALGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
DB 1 MLPWTAXGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
QY 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVASAQRMGDHTTGALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVASAQRMGDHTTGALVYA 120
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
DB 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
QY 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215
DB 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215

RESULT 8
US-09-789-561-159
Sequence 159, Application US/09789561
Patent No. US20020064818A1

GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: P2043P1
CURRENT APPLICATION NUMBER: US/09/789,561
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 159
LENGTH: 242
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-159

Query Match 50.3%; Score 1084; DB 3; Length 242;
Best Local Similarity 99.5%; Pred. No. 9.9e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLPWTALGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
DB 28 MLPWTAXGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 87

QY 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVASAQRMGDHTTGALVYA 120
DB 88 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVASAQRMGDHTTGALVYA 147
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
DB 148 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 207
QY 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215
DB 208 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 242

RESULT 9
US-10-883-936-159
Sequence 159, Application US/10883936
Publication No. US20050019866A1

GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: P2043P1
CURRENT APPLICATION NUMBER: US/10/883,936
CURRENT FILING DATE: 2004-07-06
PRIOR APPLICATION NUMBER: US/09/789,561
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 159
LENGTH: 242
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-159

Query Match 50.3%; Score 1084; DB 5; Length 242;
Best Local Similarity 99.5%; Pred. No. 9.9e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLPWTALGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
DB 28 MLPWTAXGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 87
QY 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVASAQRMGDHTTGALVYA 120
DB 88 VAPLPLGTGALRASLVHVGSRPYTEFPFGQHSSEGAADAVASAQRMGDHTTGALVYA 147
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
DB 148 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 207
QY 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215
DB 208 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 242

RESULT 10


```
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2
```

```
Query Match          42.4%; Score 913; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 2.3e-69;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 32 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
DB 1 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60
QY 92 SSGEAAQADVRAAQRMGDTHTGLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 151
DB 61 SSGEAAQADVRAAQRMGDTHTGLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 120
QY 152 GPPMQLKDLGVTTFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211
DB 121 GPPMQLKDLGVTTFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
```

```
RESULT 14
US-10-699-035A-11
; Sequence 31, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: VA domain from WARP
US-10-699-035A-31
```

```
Query Match          35.8%; Score 772; DB 4; Length 181;
Best Local Similarity 83.4%; Pred. No. 2.3e-57;
Matches 151; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
```

```
QY 32 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
DB 1 QGDLLFLDSSASVSHYEFSSRVREFVQGLVATMSPGALRASLVHVGSPHTEFTFDQY 60
QY 92 SSGEAAQADVRAAQRMGDTHTGLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 151
```

```
DB 61 SSGQAIRDAIRVAPQRMGDTHTGLALVYAKQOLFPEEAGARGVGVKVLVWTTDGGSSDPV 120
QY 152 GPPMQLKDLGVTTFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211
DB 121 GPPMQLKDLGVTTFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
QY 212 A 212
DB 181 A 181
```

```
RESULT 15
US-10-699-035A-8
; Sequence 8, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-8
```

```
Query Match          35.7%; Score 768; DB 4; Length 180;
Best Local Similarity 83.3%; Pred. No. 5e-57;
Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
```

```
QY 32 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
DB 1 QGDLLFLDSSASVSHYEFSSRVREFVQGLVATMSPGALRASLVHVGSPHTEFTFDQY 60
QY 92 SSGEAAQADVRAAQRMGDTHTGLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 151
DB 61 SSGQAIRDAIRVAPQRMGDTHTGLALVYAKQOLFPEEAGARGVGVKVLVWTTDGGSSDPV 120
QY 152 GPPMQLKDLGVTTFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211
DB 121 GPPMQLKDLGVTTFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
```

```
Search completed: February 13, 2006, 13:12:40
Job time : 86.2776 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:47:59 ; Search time 64.3077 Seconds
(without alignments)
537.392 Million cell updates/sec

Title: US-10-699-035a-6

Perfect score: 2154
Sequence: 1 MFWTALGLIALSLRLAARS.....RRPPRPVPRAPPGTASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/1/iaa/5/COMB.pep:*
2: /cgn2_6/prodata/1/iaa/6/COMB.pep:*
3: /cgn2_6/prodata/1/iaa/H/COMB.pep:*
4: /cgn2_6/prodata/1/iaa/PCTUS/COMB.pep:*
5: /cgn2_6/prodata/1/iaa/RE/COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384	17.8	77	2	US-09-513-999C-6277 Sequence 6277, Ap
2	264	12.3	755	2	US-09-919-497-57 Sequence 57, Appl
3	259	12.0	584	2	US-09-949-016-10340 Sequence 10340, A
4	259	12.0	584	2	US-09-949-016-10341 Sequence 10341, A
5	255	11.8	963	2	US-09-949-016-11519 Sequence 11519, A
6	255	11.8	963	2	US-09-949-016-11520 Sequence 11520, A
7	253.5	11.8	638	1	US-08-897-443-1 Sequence 1, Appl1
8	253.5	11.8	915	2	US-09-907-794A-34 Sequence 34, Appl
9	253.5	11.8	915	2	US-09-905-125A-34 Sequence 34, Appl
10	253.5	11.8	915	2	US-09-902-775A-34 Sequence 34, Appl
11	253.5	11.8	915	2	US-09-906-700-34 Sequence 34, Appl
12	253.5	11.8	915	2	US-09-903-603A-34 Sequence 34, Appl
13	253.5	11.8	915	2	US-09-904-920A-34 Sequence 34, Appl
14	253.5	11.8	915	2	US-09-909-064-34 Sequence 34, Appl
15	253.5	11.8	915	2	US-09-905-381A-34 Sequence 34, Appl
16	253.5	11.8	915	2	US-09-906-618-34 Sequence 34, Appl
17	253.5	11.8	915	2	US-09-906-646-34 Sequence 34, Appl
18	253.5	11.8	915	2	US-09-904-462-34 Sequence 34, Appl
19	253.5	11.8	915	2	US-09-902-736A-34 Sequence 34, Appl
20	253.5	11.8	915	2	US-09-906-722A-34 Sequence 34, Appl
21	253.5	11.8	956	2	US-09-949-016-6215 Sequence 6215, Ap
22	245	11.4	956	1	US-08-897-443-3 Sequence 3, Appl1
23	237	11.0	452	2	US-09-914-259-34 Sequence 34, Appl
24	233.5	10.8	496	1	US-08-462-180-37 Sequence 37, Appl
25	233.5	10.8	496	1	US-08-463-180-37 Sequence 37, Appl
26	233.5	10.8	496	1	US-08-001-078A-1 Sequence 1, Appl1
27	233.5	10.8	496	1	US-08-897-443-4 Sequence 4, Appl1

28	233.5	10.8	496	1	US-08-463-218-1	Sequence 1, Appl1
29	233.5	10.8	496	2	US-09-949-016-11306	Sequence 11306, A
30	233.5	10.8	496	4	PCT-US94-00253-1	Sequence 1, Appl1
31	232	10.8	405	2	US-09-312-283C-374	Sequence 374, Appl
32	228.5	10.6	481	2	US-09-914-259-36	Sequence 36, Appl
33	223.5	10.4	1036	2	US-10-104-047-2812	Sequence 2812, Ap
34	221	10.3	486	2	US-09-914-259-35	Sequence 35, Appl
35	221	10.3	486	2	US-09-976-594-278	Sequence 278, Appl
36	221	10.3	486	2	US-09-949-016-6216	Sequence 6216, Ap
37	221	10.3	507	2	US-09-949-016-9878	Sequence 9878, Ap
38	212.5	9.9	954	2	US-09-996-611D-1	Sequence 1, Appl1
39	202	9.4	1180	2	US-09-000-004A-2	Sequence 2, Appl1
40	198.5	9.2	171	2	US-09-996-611D-2	Sequence 2, Appl1
41	194	9.0	214	2	US-10-061-658-6	Sequence 6, Appl1
42	194	9.0	214	2	US-10-061-658-9	Sequence 9, Appl1
43	194	9.0	550	2	US-09-907-794A-227	Sequence 227, App
44	194	9.0	550	2	US-09-905-125A-227	Sequence 227, App
45	194	9.0	550	2	US-09-902-775A-227	Sequence 227, App

ALIGNMENTS

```
RESULT 1
US-09-513-999C-6277
; Sequence 6277, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumae Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59. US2. REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent. pm
; SEQ ID NO 6277
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 66
; OTHER INFORMATION: Xaa=Cys or Gly
US-09-513-999C-6277

Query Match          17.8% Score 384; DB 2; Length 77;
Best Local Similarity 98.7%; Pred. No. 1.7e-26;
Matches 76; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      108 MGDTHGLALVYAKQLFAEASGAPGVKVLVWTTGGSSDPVGPMDLQDLGVTVRI 167
      1 MGDTHGLALVYAKQLFAEASGAPGVKVLVWTTGGSSDPVGPMDLQDLGVTVRI 60
      61 VSTGRXNPLELSAASA 77

DB

QY      168 VSTGRXNPLELSAASA 184
      61 VSTGRXNPLELSAASA 77

RESULT 2
US-09-919-497-57
; Sequence 57, Application US/09919497
; Patent No. 673883
; GENERAL INFORMATION:
; APPLICANT: Muller, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
```

; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-57

Query Match 12.3%; Score 264; DB 2; Length 755;
Best Local Similarity 29.5%; Pred. No. 1.5e-14;
Matches 62; Conservative 44; Mismatches 90; Indels 14; Gaps 3;

QY 30 AERGLMFLDSSASVSHYEFSEFVQGLVAPLPLGTGLARSLVHGSRPYTEPPF 88
DB 3 AAKADLVFVVDGSMGIDENFNKIIISFLYSTGALINKICTDGTQVAMVQFTDDPRTFRL 62
QY 89 GQHSAGEAQAQVRAQAQMGDTHTGLALVYAKEQLFAEASGARPGVPLVWVTGSSS 148
DB 63 NAYKTEETLLDAIKIISYKGNKTKGKAIVYRDTLFTAESGTRRGIPKVIIVITDGRSQ 122
QY 149 DPVGPMEKLDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFDV-----VDDLHI 200
DB 123 DVNKRISREMOQDGYSTIFAIYADADYSELVSIKSPSAHVFPVDDPDPAFKKIEDLIT 182
QY 201 IVQELRGSLIDAMPQOLHATEITSSGFRLL 230
DB 183 FVCEHTASATCPV-----VHKDGLDLAIGFPM 207

RESULT 3
US-09-949-016-10340
; Sequence 10340, Application US/09949016

; PATENT INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10340
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10340

Query Match 12.0%; Score 259; DB 2; Length 584;
Best Local Similarity 37.2%; Pred. No. 2.9e-14;
Matches 67; Conservative 25; Mismatches 84; Indels 4; Gaps 2;

QY 34 DIMFLDSSASVSHYEFSEFVQGLVAPLPLGTGLARSLVHGSRPYTEPPGQHS 93
DB 348 DLVLLVDSKSVRPQNFELVYKRFVNIQVDFLDVSPGRTGVGVQFESSVRTEFPLGRYGT 407
QY 94 GEAADAVASAGORMDHTHTGLALVYAKEQLFAEASGARP---GVPKLVWVTGSSSDP 150
DB 408 AAEVQAQVLAVERMERTGTGLALRMVHSHSEAGAPRALNVPRGVLTDDGSSQDD 467
QY 151 VGPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGI 209
DB 468 ISVMAAKAKREGIWMYAVGVGAVERAELEIASPEALHVSVAPEFGTMTHTLENLKRSI 527

RESULT 4
US-09-949-016-10341
; Sequence 10341, Application US/09949016
; PATENT INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10341
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10341

Query Match 12.0%; Score 259; DB 2; Length 584;
Best Local Similarity 37.2%; Pred. No. 2.9e-14;
Matches 67; Conservative 25; Mismatches 84; Indels 4; Gaps 2;

QY 34 DIMFLDSSASVSHYEFSEFVQGLVAPLPLGTGLARSLVHGSRPYTEPPGQHS 93
DB 348 DLVLLVDSKSVRPQNFELVYKRFVNIQVDFLDVSPGRTGVGVQFESSVRTEFPLGRYGT 407
QY 94 GEAADAVASAGORMDHTHTGLALVYAKEQLFAEASGARP---GVPKLVWVTGSSSDP 150
DB 408 AAEVQAQVLAVERMERTGTGLALRMVHSHSEAGAPRALNVPRGVLTDDGSSQDD 467
QY 151 VGPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGI 209
DB 468 ISVMAAKAKREGIWMYAVGVGAVERAELEIASPEALHVSVAPEFGTMTHTLENLKRSI 527

RESULT 5
US-09-949-016-11519
; Sequence 11519, Application US/09949016
; PATENT INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11519
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11519

Query Match 11.8%; Score 255; DB 2; Length 963;
Best Local Similarity 27.9%; Pred. No. 1.3e-13;
Matches 68; Conservative 43; Mismatches 89; Indels 44; Gaps 4;

QY 5 TALGLALSURL-----ALAKGARGPPASA----- 30

Db 32 TALPLLLALMKEMKLAGCFLLLIGQIVLPAEARESRGRSISRGHARTHPOTALLSS 91
QY 31 ---PRGDMFLDSSASVSHYFSRVEPVGOVAPLPLGTALRASLVHVGSRPYTEP 87
Db 92 CENKRADLVFIIDSSRVTHDYAKVEFIVDILQFLDGPVTRVGLLQYGSTVKNES 151
QY 88 FCGHSSGEAQAODVRAAQMCDTHGTALVYAKEOLFAEASGARP---GVPEVLVWYTD 144
Db 152 LKTFKRKSEVERAVKMRHLSTGTMTGLAQYALNIAFSEBEGARPLRENVPRVIMIVTD 211
QY 145 GGSDDPVGPMEOLKDLGVTVFIVSTGRGNFLELSAASAPAEKHL---HFVDDVLDHI 200
Db 212 GRPQDSVAEVAAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVLVANFSQIETLIS 271
QY 201 IVOE 204
Db 272 VFQK 275

RESULT 6
US-09-949-016-11520
; Sequence 11520, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11520
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11520

Query Match 11.8%; Score 255; DB 2; Length 963;
Best Local Similarity 27.9%; Pred. No. 1,3e-13;
Matches 68; Conservative 43; Mismatches 89; Indels 44; Gaps 4;

QY 5 TALGALSLRL-----ALARSGAERGPPASA----- 30
Db 32 TALPLLLALMKEMKLAGCFLLLIGQIVLPAEARESRGRSISRGHARTHPOTALLSS 91
QY 31 ---PRGDMFLDSSASVSHYFSRVEPVGOVAPLPLGTALRASLVHVGSRPYTEP 87
Db 92 CENKRADLVFIIDSSRVTHDYAKVEFIVDILQFLDGPVTRVGLLQYGSTVKNES 151
QY 88 FCGHSSGEAQAODVRAAQMCDTHGTALVYAKEOLFAEASGARP---GVPEVLVWYTD 144
Db 152 LKTFKRKSEVERAVKMRHLSTGTMTGLAQYALNIAFSEBEGARPLRENVPRVIMIVTD 211
QY 145 GGSDDPVGPMEOLKDLGVTVFIVSTGRGNFLELSAASAPAEKHL---HFVDDVLDHI 200
Db 212 GRPQDSVAEVAAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVLVANFSQIETLIS 271
QY 201 IVOE 204
Db 272 VFQK 275

RESULT 7
US-08-897-443-1
; Sequence 1, Application US/08897443
; Patent No. 5981263

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Kaser, Mathew
; TITLE OF INVENTION: HUMAN MATRILIN-3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,443
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0348 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UTRSN02
; CLONE: 681719
US-08-897-443-1

Query Match 11.8%; Score 253.5; DB 1; Length 638;
Best Local Similarity 31.6%; Pred. No. 9.9e-14;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERGPPASAPRGDMFLDSSASVSHYFSRVEPVGOVAPLPLGTALRASL 75
Db 45 ALLESSCEN-----KRALVFIIDSSRVTHDYAKVEFIVDILQFLDGPVTRVGL 98
QY 76 VHVGSRPYEPFCGHSSEAAQDAVRAAQMCDTHGTALVYAKEOLFAEASGARP-- 133
Db 99 LQYGSTVKNESLKTFFKRSEVERAVKMRHLSTGTMTGLAQYALNIAFSEBEGARPLR 158
QY 134 -GVPEVLVWYTDGGSDDPVGPMEOLKDLGVTVFIVSTGRGNFLELSAASAPAEKHL-- 190
Db 159 ENVPVIMIVTQGRPQDSVAEVAAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVL 218
QY 191 --HFVDDVLDHIIVOE 204
Db 219 VANFSQIETLISVFQK 234

RESULT 8
US-09-907-794A-34
; Sequence 34, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kijavlin, Jennie P.
APPLICANT: Mathier, Jemie P.
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,794A
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-794A-34
Query Match 11.8%; Score 253.5; DB 2; length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;
QY 16 ALARGAERGPAPASDRGLMLLDSASVSHYERFRVREFVQGLVAPLPLGTGLRSL 75

Db 45 ALLESSCEN-----KRADLVFTIDSSRVNTHDAKVEKFLVDLQFLDGPDPVTRGL 98
QY 76 VHVGSRPYTFEPFGQSSGGEAAQDAVRASAGRMGTHGLALVYAKEOLFAEASGARP-- 133
Db 99 LQYSTVNEFSLKTFKRKSEVERAVKMRHLSITGTMTGLAQYMLNLAFSABGARPLR 158
QY 134 -GVPRVLVWVTDDGSSDPVGPMEQLKDLGTVTFIVSTGRGNFLELSAASAPAEKHL-- 190
Db 159 ENVPFRVIMIVTDGRQDSVAEVAARAKBDTGILIFAIGQVDFTNLSIGSEPHEDHVF 218
QY 191 --HFVDVDDHIIYOE 204
Db 219 VANFSQIETLTTSVFOK 234
RESULT 9
US-09-905-125A-34
Sequence 34, Application US/09905125A
Patent No. 6664376
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mathier, Jennie P.
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905,125A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30

```

; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-125A-34

Query Match      11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERPPASARGDLMFLDSSASVSHYEFKRVREFVQGLVAPLPGLGALRASL 75
DB 45 ALLESSCEN-----KRADLVFIIDSSRSVNTHDYAKVEFVDDIQLFDIGDVTYRVL 98
QY 76 VHVGSRPYTEPPFGQSSGEAAODAVASAOBMGDTHTGLALVYAKEQLFAEASGARP-- 133
DB 99 LQYGSTVKNKEFSLKTRKSEVERAVKRMHLSGTMTGLAIQYALNIAFSEAEGRPLR 158
QY 134 -GVPKYLVWTDGSSDPVGPMPQELKDLGVTFVIVSTGRGNFLELSAASAPAEKRL-- 190
DB 159 ENVPRIYIMVTGDRPDQSVAEVAAKARDTGILIFALGVGVDPNTLKSIGSEPHEDHVF 218
QY 191 --HFVDVDDLHIIVQE 204
DB 219 VANFSQIETLTSTVFOK 234

RESULT 10
US-09-902-775A-34
; Sequence 34, Application US/09902775A
; Patent No. 6686451
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Macher, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Peoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
```

```

; CURRENT APPLICATION NUMBER: US/09/902,775A
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-775A-34

Query Match      11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERPPASARGDLMFLDSSASVSHYEFKRVREFVQGLVAPLPGLGALRASL 75
DB 45 ALLESSCEN-----KRADLVFIIDSSRSVNTHDYAKVEFVDDIQLFDIGDVTYRVL 98
QY 76 VHVGSRPYTEPPFGQSSGEAAODAVASAOBMGDTHTGLALVYAKEQLFAEASGARP-- 133
DB 99 LQYGSTVKNKEFSLKTRKSEVERAVKRMHLSGTMTGLAIQYALNIAFSEAEGRPLR 158
QY 134 -GVPKYLVWTDGSSDPVGPMPQELKDLGVTFVIVSTGRGNFLELSAASAPAEKRL-- 190
DB 159 ENVPRIYIMVTGDRPDQSVAEVAAKARDTGILIFALGVGVDPNTLKSIGSEPHEDHVF 218
QY 191 --HFVDVDDLHIIVQE 204
DB 219 VANFSQIETLTSTVFOK 234

RESULT 11
US-09-906-700-34
; Sequence 34, Application US/09906700
; Patent No. 6723535
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Db      45  ALLESCEN-----KRALVITIISSSKSVNTHDPAKKEITVILLOFLDGPVTVVGL 98
Oy      76  VHWGSRAPTEFPFGHSSGEAAQDAVRASAOQMGDHTHTGLAVYAKQQLFNAESGAP-- 133
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      99  LQGSTVTKVIEFSLKTFKRKSEVERAVKMRHLSGTMTGTGLAIQYALNIAFSEAGAPPLR 158
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Oy      134 -GVPKILVWYTTDGGSSDPFGPQMOKLDGTVVFVTSVGRGNFLELSAASAPAEKFL-- 190
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      159 ENNPRVIMTITDPRPDSDVAEVAARAKRDTGLIFAIQGVDFNTLKSIGSEPEHDVFL 218
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Oy      191 --HFVVDLHIVQE 204
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      219 VANFSQIETITSVFOK 234

RESULT 12
US-09-903-603A-34
/ Sequence 34: Application US/09903603A
/ Patent No. 6767995
/ GENERAL INFORMATION:
/ APPLICANT: Genentech, Inc.
/ APPLICANT: Aeshkenazi, Avi
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Baton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gertlesen, Mary E.
/ APPLICANT: Goddard, A.
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Kijavlin, Ivar J.
/ APPLICANT: Hillan, Kenneth, J.
/ APPLICANT: Mathier, Jennie P.
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ TITLE OF INVENTION: Acids Encoding the Same
/ FILE REFERENCE: GNE.1618P2C12
/ CURRENT APPLICATION NUMBER: US/09/903,603A
/ CURRENT FILING DATE: 2001-07-11
/ PRIOR APPLICATION NUMBER: PCT/US00/04414
/ PRIOR FILING DATE: 2000-02-22
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564

```


APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Goddard, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/909,064
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-909-064-34
Query Match 11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;
QY 16 ALARSAERPPAPRSGDMLPDDSGASVSHFEFVRFVQGVAPLPLGAGAPASL 75
DB 45 ALLESCEN-----KRAADVFTIIDSRSVNTHDYAKVKEFIVDILQFLDGPVTRVGL 98

QY 76 VHGSRPYTERPPFGCHSGEAAQDAVRASAOQMDTHTGLALVYAKEQLFAEASGAP-- 133
DB 99 LQYGSTVAKNEFSKLTKFKKSEVERAVKMRHLSTGTMTGLAIQYALNTAFSEAEQAPLR 158
QY 134 -GVPRVLTWMTDGGSDPVGPPMOELKDLGVTFIVSTGRGNPFLSAASAPAKHL-- 190
DB 159 ENVPVIMIVTDGRPODSVAEVAAMAKRDTGLITPAIGVGQVDFNTLKSIGSEPHEDHVF 218
QY 191 --HPVDVDDLHIIVQE 204
DB 219 VANFSQIETLTSVFQX 234
RESULT 15
US-09-905-381A-34
Sequence 34, Application US/09905381A
Patent No. 6818746
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Goddard, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905,381A
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02

```

? PRIOR APPLICATION NUMBER: PCT/US99/28565
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US99/30095
? PRIOR FILING DATE: 1999-12-16
? PRIOR APPLICATION NUMBER: PCT/US99/30911
? PRIOR FILING DATE: 1999-12-20
? PRIOR APPLICATION NUMBER: PCT/US99/30999
? PRIOR FILING DATE: 1999-12-20
? PRIOR APPLICATION NUMBER: PCT/US00/00219
? PRIOR FILING DATE: 2000-01-05
? NUMBER OF SEQ ID NOS: 423
? SEQ ID NO 34
? LENGTH: 915
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-905-381A-34

```

Query Match	11.8%;	Score 253.5;	DB 2;	Length 915;
-------------	--------	--------------	-------	-------------

```

QY      76 VHVGSRPYTEFPFGHSGEAAODA VRASAO RMGDTHTGIALVYAKEOLF AEA SGARP -- 133
Db      99 LQYGVSTVKNQEFSLKTPFKRSVEVERAVKRM RHLSTGMTGIALGYALNIAFSEAEGRPLR 158
QY      134 -GVPKVLVWYDTGGSSDPVGPMMOEKLDI GVTVFIVYSTGRGNFLELSAAASAPAEKHL -- 190
Db      159 ENVPPEVIMIVTDEGRPODSVAEVAAKARDTGIL IFAIGVQVDVFNTLKSIGSEPHEDHVEL 218
QY      191 --HFVDVDDLHIIVOE 204
Db      219 VANFSQIETLTLSVFOK 234

```

Search completed: February 13, 2006, 07:50:52
Job time : 65.3077 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:39:36 ; Search time 9.24221 Seconds
(without alignments)
2243.517 Million cell updates/sec

Title: US-10-699-035A-5

Perfect score: 2380
Sequence: 1 atgctccctggacgagcgct.....ccgcacccgctgacgcgttaa 1254

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-G/abses/ABSGWB.spool/US10699035/runat 13022006 062447 25513/app.query.fasta_1
-DB=Issued_Patents_AA -QPM=fasten -SUFFIX=ra1 -TINMATCH=0.1 -LOOFC=0
-LOOEXT=0 -ONITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs02p -USER=US10699035 @CGN 1.1 101 @runat 13022006 062447 25513
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPLLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FCAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued Patents AA:*
2: /cgnt2_6/prodata/1/iaa/5/COMB.pep:*
3: /cgnt2_6/prodata/1/iaa/H/COMB.pep:*
4: /cgnt2_6/prodata/1/iaa/PC/US.COMB.pep:*
5: /cgnt2_6/prodata/1/iaa/BB.COMB.pep:*
6: /cgnt2_6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384	16.1	77	2	US-09-513-999C-6277, AP Sequence 6277, AP
2	287.5	12.1	860	2	US-09-252-991A-28607, A Sequence 28607, A
3	281.5	11.8	663	2	US-09-252-991A-10843, A Sequence 10843, A
4	279	11.7	511	2	US-09-252-991A-26078, A Sequence 26078, A
5	279	11.7	2294	2	US-09-252-991A-11731, A Sequence 11731, A
6	275.5	11.6	917	2	US-09-252-991A-25101, A Sequence 25101, A
7	272	11.4	650	2	US-09-252-991A-20375, A Sequence 20375, A
8	270	11.3	492	2	US-09-252-991A-28339, A Sequence 28339, A
9	269.5	11.3	689	2	US-09-252-991A-31790, A Sequence 31790, A
10	269	11.3	1706	2	US-09-252-991A-31760, A Sequence 31760, A
11	268	11.3	663	2	US-09-252-991A-30843, A Sequence 30843, A
12	267	11.2	638	2	US-09-252-991A-27068, A Sequence 27068, A

13	266	11.2	755	2	US-09-919-497-57	Sequence 57, Appl
14	264.5	11.1	783	2	US-09-252-991A-18035	Sequence 18035, A
15	263.5	11.1	590	2	US-09-252-991A-19127	Sequence 19127, A
16	261	11.0	369	2	US-09-252-991A-25394, A	Sequence 25394, A
17	259.5	10.9	642	2	US-09-252-991A-24767, A	Sequence 24767, A
18	259	10.9	584	2	US-09-949-016-10340, A	Sequence 10340, A
19	259	10.9	584	2	US-09-949-016-10341, A	Sequence 10341, A
20	259	10.9	819	2	US-09-252-991A-19569, A	Sequence 19569, A
21	258	10.8	808	2	US-09-252-991A-32826, A	Sequence 32826, A
22	257	10.8	1228	2	US-09-252-991A-17764, A	Sequence 17764, A
23	256	10.8	2294	2	US-09-252-991A-17231, A	Sequence 17231, A
24	255.5	10.7	302	2	US-09-252-991A-19529, A	Sequence 19529, A
25	255	10.7	963	2	US-09-949-016-11519, A	Sequence 11519, A
26	255	10.7	963	2	US-09-949-016-11520, A	Sequence 11520, A
27	253.5	10.7	638	1	US-08-897-443-1, A	Sequence 1, Appl1
28	253.5	10.7	770	2	US-09-252-991A-30323, A	Sequence 30323, A
29	253.5	10.7	915	2	US-09-907-794A-34, A	Sequence 34, Appl
30	253.5	10.7	915	2	US-09-905-125A-34, A	Sequence 34, Appl
31	253.5	10.7	915	2	US-09-902-775A-34, A	Sequence 34, Appl
32	253.5	10.7	915	2	US-09-906-700-34, A	Sequence 34, Appl
33	253.5	10.7	915	2	US-09-903-603A-34, A	Sequence 34, Appl
34	253.5	10.7	915	2	US-09-904-920A-34, A	Sequence 34, Appl
35	253.5	10.7	915	2	US-09-909-064-34, A	Sequence 34, Appl
36	253.5	10.7	915	2	US-09-905-381A-34, A	Sequence 34, Appl
37	253.5	10.7	915	2	US-09-906-618-34, A	Sequence 34, Appl
38	253.5	10.7	915	2	US-09-906-646-34, A	Sequence 34, Appl
39	253.5	10.7	915	2	US-09-904-462-34, A	Sequence 34, Appl
40	253.5	10.7	915	2	US-09-902-722A-34, A	Sequence 34, Appl
41	253.5	10.7	915	2	US-09-906-726A-34, A	Sequence 34, Appl
42	253.5	10.7	956	2	US-09-949-016-6215, A	Sequence 6215, Ap
43	253	10.6	405	2	US-09-312-283C-374, A	Sequence 374, App
44	252.5	10.6	681	2	US-09-252-991A-24567, A	Sequence 24567, A
45	252.5	10.6	1065	1	US-08-642-255-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-09-513-999C-6277
; Patent 6277, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Maline Edwards, J.B.
; APPLICANT: Duclet, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 6277
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 66
; OTHER INFORMATION: Xaa=Cys or Gly
US-09-513-999C-6277
Alignment Scores: 1,61e-18
Pred. No.: 384.00
Score: 98.74
Percent Similarity: 98.74
Best Local Similarity: 16.14
Query Match: 2
DB: 0
US-10-699-035A-5 (1-1254) x US-09-513-999C-6277 (1-77)

RESULT 6
US-09-252-991A-25101
; Sequence 25101, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25101
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25101

Alignment Scores:
Pred. No.: 6,06e-11 Length: 917
Score: 275.50 Matches: 165
Percent Similarity: 31.9% Conservative: 21
Best Local Similarity: 28.3% Mismatches: 198
Query Match: 11.6% Indels: 199
DB: 2 Gaps: 31

US-10-699-035A-5 (1-1254) x US-09-252-991A-25101 (1-917)

QY 3 GCTCCCTGAGACGCGCTCGGCTCTGAGCTGCGGCTGCGCGG 56
DB 362 AAlaProCyAlaGlyProSerProAlaArgGlyAlaArgAlaAlaGlyAlaSerPro 381
QY 57 GAGCGCGCGAGCGCGCTCCACGACATCAAGCCCCGAGGCGACCTGATGTTCTGCT 116
DB 382 AAspArgAlaGlnAlaArgAlaThrAsp-----GlyGlyProGlnAlaProAla 397
QY 117 GGAACAGCTCAAGCCAGCT-----CTTCACATA 143
DB 398 ProAlaAlaArgProAlaArgAspValGlyGlyLeuArgProAlaAlaThrLeuArgAla 417
QY 144 CGAGTTCTCCGCGGTTGCG-----GGCTCCACTGCCCCCTGG 161
DB 418 ArgArgArgProGlyGlyAlaArgGlyAlaSerProGlnSerArgProAlaArgGlyAspPro 437
QY 162 -----GAGATTGTGGGCAAGCTGCT---GGCTCCACTGCCCCCTGG 200
DB 438 ArgArgCyAlaArgProGlyAlaArgAlaProAlaThrGlySerGlyAlaArgAlaProGly 457
QY 201 CACCGCGCGCGCTCGCGCCAGCTGTCGACGTCGGGCACTCGGCCATACCGAAGTTCCTCC 260
DB 458 ArgArgSer-----GlyGlnGlyAlaProGlyGlnGlyGlyGlyAlaArgArgPro 474
QY 261 CTT-----CGGCGACGACAGCTCGGCTGAGGCTGCGCCAGATGCGCT 302
DB 475 AlaValArgAspArgGlnAlaArgProAlaArgAlaGlnAlaSerProGlnAlaSerGly 494
QY 303 GCGTCCTTGTCCCGACCGCATGGGTGACACCCACTGCGCTGGCGCTGCTGATGCCAA 362
DB 495 AlaLeuArgGlnProAlaAspArg---ArgGlnAlaPhePro----- 507
QY 363 GGAACAGCTGTTGCTGGAAGCATCAGTGCCTGGCGAGGGGTGCCAAAGTGTGCTGCT--- 419
DB 508 -----LeuGlnArgValProThrGlnGlyArgProGlyAlaGlyGly 521
QY 420 -----GTGGTGACAGATGGCGGCTCCAG 443
DB 522 AlaAspArgProValSerAlaGlnAlaArgTyrProLeuSerThrAspArgProGlyLeuAla 541

QY 444 CGACCTGCT-----GGGCCCCCATGCGAGAGCTCAAGCA----- 479
DB 542 ArgArgCyAlaArgGlyAspProGlyAlaProAlaAlaAlaArgAlaGlyLeuAlaArg 561
QY 480 -----CGCGGCGTCAAGCTGTTGATTTGACAGCAAGCGCGCAAGCACTTCTCT 527
DB 562 ArgLeuAlaProProGlyAlaArgAlaGlyAlaSerGlyAlaSerArgGlySerProAlaPro 581
QY 528 GGA---GCTGTCAAGCGCTGCTGAGCCCTCGCCAGAGAACCTGCACTTTGTGACGCT 584
DB 582 GAlaArgAlaSerArgProAspGlnProAlaAlaGlyLeuProPheLeuValGlyAla 601
QY 585 GGAATCACTGCACATATTG----- 605
DB 602 -----ThrAlaGlyArgCyAspSerLeuArgGlnArgLeuAlaTyrGlnLeuVal 619
QY 606 -----CSAAGAGCTGAGGGGCTGATTCGAGTGGAGGCGCGAGAGCTGCA 653
DB 620 ArgProAlaGlnAlaProGly---HisAlaArgLeuAlaValGlyGlnAla--- 637
QY 654 TGCCACGAGATCAAGTCCAGCG----- 677
DB 638 -----GlyAspRnGlyGlnArgAlaLeuArgRnAlaArgArgGlnAlaArgLeuSer 655
QY 677 ----- 677
DB 656 AlaAlaAlaGlyAlaSerProGlyAlaArgArgArgRnAlaAspGlyAlaArgGlnArg 675
QY 678 -----CTTCGCGCTGAGCTGAGCC-----ACCCCT 701
DB 676 ThrGlnAlaArgAlaAlaValLeuAlaLeuGlyLeuAlaArgLeuAlaArgAlaGly 695
QY 702 GCTGACCGCAAG-----CTCGAGCTATGATGCTGGA-----GCT 737
DB 696 AlaGlnGlnAlaArgSerGlnProGlyAspLeuGlyAlaAlaCyAspProGlyAlaArgArgPro 715
QY 738 GGTGCCCGAGCGCGCGCGGCTGCAAGACCGCAGAGAGCTGCCAGGAAAGCCAGCGGA 797
DB 716 GlyValGlnProGlyAlaGlySerAspArgPheProLeuArgAlaLeuCyAsGlyHisAla 735
QY 798 CTGAGTCTGGGCGCGGCTCGAACCC-----GGAACAGGA 830
DB 736 ArgLeuGlnGlyAlaSerProGlyAlaSerProGlyAlaSerArgValLeuGlyAlaArgGly 755
QY 831 CTACGACGCT-----GGCGCTAGTGCCTGAGTCCACAGCTGCGGCTTGAAGCGCCCA 881
DB 756 LeuArgArgArgProProGlyAlaAlaAlaGlyGlyGlnHisArg-----ProGlnArgAla 773
QY 882 GATCTGCGGGGTGCGCAGCGCGCCAGAGAGCGCGGCGCAGAGCATGTCATCTCCCA 941
DB 774 AlaAlaAlaAlaAlaGlnPheArgAlaGlyGlnAlaProAlaThrArgRnAlaAla 793
QY 942 ---GGCGCGCGCGCG-----CAGCTCCGCGGTGAGTTGGGCGCCAGCGCTGGG 986
DB 794 GlyArgSerGlnAlaLeuGlyGlyAlaSerSerProGlnArgProAlaGlyProAlaArgAlaAla 813
QY 987 CTCAGCGCGCGG---GCTCGAGCTACACAGTGCAGTTCGGGCGCGCTGCGGCGGAGAGC 1043
DB 814 LeuSerLeuValAlaValAlaAlaLeuSerArgArgGlyArgGlyAlaThrGlyArg---Pro 832
QY 1044 GCAAGCGGTGAGAGTCCCGCGCGCGCGCAGACAGTGCACACAGCTGCGAGGCGCTGCGCGCGG 1103
DB 833 ValArgGlyGlyGlyAlaArgSerAlaProAlaAlaArgArgArgSerGlyAlaGlyAlaAsp 852
QY 1104 CACGCGCTACCTGAGTGAACCGTGAACCGCGCGCTT---CCGCTCGGCGCGCGAGAGCGCGCT 1160
DB 853 ArgArgSerProGlyAlaArgPheProArgSerValaArgProAspGlyProAlaGlyGlySer 872
QY 1161 GTCCGCAAGGCTGCGACGCGGA----- 1184
DB 873 AlaArgArgGlyAlaAspArgAlaArgArgSerAlaArgArgGlnArgArgProAlaValaArgSer 892
QY 1185 -----CGGCGCGCGCGCGCGCGCCAGCGCGCGCTGCGCGCGCGAGACCGCGGAGAC 1235


```

Oy      1152 GAGGGCGCTGTCCGCGAAGGCTGCAC---GCCGACGGCCCCGGCGCCGACGCC 1208
           |||
Db      516 HisArg-----ArgProGlyGlnArgProGlnArgGlnProThrProAlaLeu 533
           |||
Oy      1209 CGTGGCCCCGGCGCCGACCCCGGGGAC 1235
           ::|||::|||::|
Db      534 CysLeuArgGlnProGlnProGlyGln 542

RESULT 8
US-09-252-991A-28339
; Sequence 28339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28339
; LENGTH: 492
; TYPE: PRt
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28339

```

```

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31790
; LENGTH: 689
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31790

Alignment Scores:
Pred. No.: 1,47e-10 Length: 689
Score: 269.50 Matches: 159
Percent Similarity: 31.5% Conservative: 14
Best Local Similarity: 28.9% Mismatches: 158
Query Match: 11.3% Indels: 219
DB: Gaps: 33

US-10-699-035a-5 (1-1254) x US-09-252-991A-31790 (1-689)
QY 12 GAGCGGCTGCGG-----CCTGGCCCTGAGCTTGCGGCTGCGGCTGCGGCG 56
DB 194 AaPGLYAlaArgAlaAlaLeuArgAlaAaPGLYArgArgLeuSerProAlaArgAlaArgLeu 213
QY 57 GAGCGGCGGAGAGCGGCTCCACCAACATCAGCCCGCGAGGGA----- 101
DB 214 ArgArgGlnGlyHisArgSerProProAlaArgLeuProAlaGlyArgThrGlnAlaGlu 233
QY 102 CCTGATGTTCTT---CTGGACAGCTCAGCCGCTCTCTACACTAGATTCTCCGGGT 158
DB 234 ProAaPLeuProGlyAlaAaPHisHisArgGlnArg----- 245
QY 159 TCGGGAGTTTGGGGGCGAGCT-----GGTGGCTCCAGTCCCTGGGACCGGGGCGCT 212
DB 246 GlyAlaValAlaGlyAaPHisArgArgGlyAlaGlyArgAlaProGlyArgProGlyLeu 265
QY 213 GCGTGCAGCTGTGTGCACGTGGGCGAGTGGCCATACACCGATTCCCTTGGCGCAGCA 272
DB 266 GlyAlaArgLeuGly----- 270
QY 273 CAGCTCGGCTGAGGCTGCCCAAGATCGGCTGCTTCTGCCAAGCATGGGTGACAC 332
DB 271 GlnLeuAlaArgGlyAaPProArg-----ProCysHisGly--His 283
QY 333 CGACAC----- 338
DB 284 ProHisGlyAaPGLYGLYGLYArgProAlaValLeuGlnGlyArgArgHisArgProAaP 303
QY 339 -----TGG---CCTGGCCGCTGCTATGCCAA-----GGAACAGCTGT 374
DB 304 ArgArgGlyLeuProArgArgGlyLeuHisGlnProLeuArgArgAlaGlyAlaProVal 323
QY 375 TCGTGAAGCATGAGTGGCCCGGCGAGGGGTGCCAAAGTCTGTGTGGGTGACAGATGG 434
DB 324 GlnArgArgGlyArgArgGlyAlaArgAlaGlnGlnGlyProGlyGlyProAlaArgAla 343
QY 435 CGGCTCCAGCAACCTGTGGGCCCCCGCATGCAAGAGCTCAAGAGCTGGGCGTCAACCT 494
DB 344 AlaLeuArgAaPProArgArgProProHisAlaAlaLeu-----GlyGlnHisArg 360
QY 495 GTTCAATTGACAGACCGCGGAGCAACTTCTGTGAGCTGTGACCGCTGCTCAGCCCC 554
DB 361 ProAlaArgArgCysArgProArg-----ProGlyAlaTyrArgArg----- 374
QY 555 TGCAGGAAGACCTGCACTTGTGAGCTGTGATGACTTGACATCAT----- 602
```

```

DB 375 -----GlyArgGly--ProValProHisArgSerAlaVal 385
QY 603 -----TGTCCAGAGCTAGAGGCTCATCTT-----CGCATGCG 638
DB 386 HisAaPArgArgProLeuProGlnArgGlnGlyThrAlaGlyAaPLeuProArgAlaAla 405
QY 639 GCCGAGCAGCTCCATGCCACGAGATCAC----- 668
DB 406 GlnCysLeuProProAlaAlaGlyAaPHisAlaHisProGlyTyrArgArgArgGlnGly 425
QY 669 -----GTCCAGCGGCTT----- 680
DB 426 AlaValLeuLeuProAaPGLYGLYArgGlnProValProAaPLeuAlaArgHisProArg 445
QY 681 -----CCGCTTGGC---CTGGCACCCCTGTGACCCGAGCATCGGCTCAATGT 728
DB 446 HisProArgProProGlyAaPLeuProGlyProAaPProArgHis----- 460
QY 729 GCTGAGCTGTGCCAGCGGCCAGCCGAGGGGTGCMAAGACCCGACGACTGCCAGGGA 788
DB 461 -----AlaGlnGlyGlnArgArgThrGlyGlnProAlaTyrProAlaAlaAaP----- 476
QY 789 CGCCACGAGACTGATCTGGGCGCGCTGACCCGGACACGAGCTAGACGTGGCGCTAGT 848
DB 477 -----AaPLeuArgHisProArgArgAlaGlyArgGly----- 486
QY 849 GCCTGATCTCAACGTGCGCTCTGTGAGCCCGCCAGATCTCT-----CGGGGTGCG 896
DB 487 -----ProAlaProAaPProProArgLeuGlnGlyArgGlyAla 498
QY 897 CAGCGGCGCAGAGAGCGCGGCGCAGACGCACTGTCAT----- 935
DB 499 ArgArgGlyArgGlyHisArgHisAlaAlaHisArgHisAaPGLYArgAaPSerArgArg 518
QY 936 ---CTCCACGCGCGCGCGCGAGCTCCGCTGAGTTGGGC----- 974
DB 519 ArgValProAaPProArgArgAlaGlyProProGlyArgLeuProPheGlyArgPheGlnArg 538
QY 975 -----CCCAGCGCTGGGCTC-----AGCCGCGGCGCTCGG 1004
DB 539 ProAaPProValProAlaGlyGlyArgArgProGlnGlnSerAlaGlyArgArgProLeuArg 558
QY 1005 CTACCAAGTGCAGTTTGGGCGCTGCGGAGCGGAGGCGCAGCGGCTGAGATGCTCCGCG 1064
DB 559 LeuProAlaSerGlyArgAlaAlaCys-----ValGlnGlnGlyGlyArgArg 574
QY 1065 GGGCGGCACTGCACACAGCT----- 1085
DB 575 CysProProGlyArgGlnAlaGlyGlnHisLeuArgArgAaPGLYArgArgSerArgGly 594
QY 1086 -----GCAAGGCGCTGGGCGCGCGGCGCCTACTGTGAGACGTGACCGCGCTTCCG 1139
DB 595 CysArgArgAlaAlaAaPGLYAaPGLYLeuArgGlnProValAaP---GlnArgHisGlnPro 613
QY 1140 CTC-----GGGCGCGAGAGCGGCTGTCCG 1166
DB 614 AlaGlnGlyGlnValAlaAlaAlaProAaPHisProGlyGlnGlyProGlyProAlaArg 633
QY 1167 CAGGCTGTCAGACCGCCGAGCGCGCGCGCGCGCCAGCGCGCGCGCGCG----- 1217
DB 634 ProValAlaHisLeuArgGlnProAlaGlyHisProGlnLeuAlaAlaProGlyValAla 653
QY 1218 GCGCCGACCGCGGGAGCGC---CAGCGG 1244
DB 654 GlnProArgProGlySerArgAaPGLNPro 663

RESULT 10
US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```


ORGANISM: Homo sapiens
US-09-919-497-57

Alignment Scores:

Pred. No.:	2	59e-10	Length:	75
Score:	26	.00	Matches:	118
Percent Similarity:	38.5%		Conservative:	78
Best Local Similarity:	23.2%		Mismatches:	169
Query Match:	11.2%		Indels:	145
DB:	2		Gaps:	18

US-10-699-035A-5 (1-1254) X US-09-919-497-57 (1-755)

Qy	88	GGCCCCAGAGGGAGCTAGATGTTCTCCGAGCAAGCTCAGCAGGGTCTTCATCAGAG	147
Db	3	AlaAlaAlaValAlaAspLeuValAlaPheMetValaLaArgLysSerTPrSerIleGlyAspGluAsn	22
Qy	148	TTCTCCCGGGTTCGGGAGATTGTGTGGGGCAGCTGGTGGCTCCACTG--CCCTGGGCAAC	204
Db	23	PheAsnLysValIleIleSerPheLeuTySerThrValGlyAlaLeuAsnLysIleGlyThr	42
Qy	205	GGGGCCCTGGTGCAGCTGTGGAGTGCACGCGGAGCATCGGCACTCCGACCAAGCTCCCTTC	264
Db	43	AspGlyThrGlnValAlaMetValaGlnPheThrAspAspProAlaGlnThrGluPheLysLeu	62
Qy	265	GGCCAGACAGACTCGGGTAGAGCTGGCCAGATGGGATGGCTGTCTTGTGCCACGGATG	324
Db	63	AsnAlaTyThrThyThrLysGlnThrLysLeuAspAlaIleLysHisIleSerTyLysGly	82
Qy	325	GGAGACACCCACCTATGGCTGGCGGCTGGCTGTATGCGACAGAAACAGCTGTTTGGCTGAACA	384
Db	83	GlyAsnThrLysThrGlyLysAlaIleLysTyValaArgAspThrLysPheThrAlaGlu	102
Qy	385	TCAGAGTCCCGGACAGAGGGTGGCCAAAGTGTGTGTGGTGGTGCACAGATAGCGGCTCCAGC	444
Db	103	SerGlyThrArgArgGlyLysLeuProLysValIleValIleThrAspGlyArgSerGln	122
Qy	445	GACCTGTGGGCCCCCAGACAGACTCAAGACCTGGGGCTGACCGTGTTCATTGTC	504
Db	123	AspAspValaAsnLysIleSerArgGlnMetGlnLeuAspGlyTySerIlePheAlaIle	142
Qy	505	AGCAGCGGCGGAGGCAACTTCCGGAGAGCTGTAGAGCGGCTGAGCGCCCTGACCGCCAGAGAG	564
Db	143	GlyValaAlaPheAlaAspTySerGlnLeuValSerIleGlySerLysProSerAlaArg	162
Qy	565	CACCTGCACCTTTGTGAC--	582
Db	163	HisValaPhePheValaAspAspPheAspAlaPheLysLysValIleGluAspGlnIleIleThr	182
Qy	583	--	GTGGATGAC 591
Db	183	PheValaCysGlnThrAlaSerAlaThrCysProValaValHisLysAspGlyIleAspLeu	202
Qy	592	CTGCACATCATTTTCCACAGAGCTGAGGGCTCCATT--	627
Db	203	AlaGlyLysPheLysMetMetGlnMetPheGlyLeuValaGlnLysAspPheSerValaGlu	222
Qy	628	--CTGCGCATGGCGCG--	CAGACAGCTCCATGCCACG 660
Db	223	GlyValaSerMetGlnProGlyThrPheAsnValaPheProCysTyGlnGlnLeuHisLysAsp	242
Qy	661	GAGATACACGTCACGCGGCTTCCGCGCTGCGCTGCGCCACCGCTGACCGGACAGAC--TGC	717
Db	243	AlaLeuValSerGlnProThrArgTyLysHisIleProGlnGlnLysProSerAspTyThr	262
Qy	718	GGCTACTATGTCTGAGCTGATGGTCCACAGCGC--	CAGCGCGGGCTGCAAGACG 717
Db	263	IleSerPheLeuPheArgLysLeuProAspAlaThrProGlnGlnProPheAla--	279
Qy	772	CAGACAGTCCAGAGAAACGCAACAGACTGGATCTGGCGCGCTGACCCGGAACACGAC	831
Db	280	--	LeuTyrGlnLysLeuHisLysAsnSerAsp 285

QY 832 AACAGCAGGGGCGTAGTGGCTCGAGTCCAACTGAGGCTCTCTGAGG-----CCC 879
 Db 290 ProLeuValIcIyValIIleLeuAspAsnGlyIyIyAsnIleuThrTyPheAsnTyAsp 309
 QY 880 CAGATCTCTGGCGGTGCGCCAGCGCGGCGAGAGAGCGCGGCGAGAG----- 924
 Db 310 GlnSerGlyAspPheGlnThrValThrPheGlu---GlyProGluIleArgIyIlePhe 328
 QY 925 -----CCGATGTCAATCTTCCACAGCCCGGCGCCGAGCCTCCG- 962
 Db 329 TyrGlySerPheHisIyIyLeuHisIleIleValIalSerIleuThrIleuValIyValIle 348
 QY 963 -----CGGATTTGGGGCCCGGCGCTGGCGCTC---AACCGCGCG 998
 Db 349 AspCysIyIyGlnValIcIyGlnIyAlaMetAsnAlaSerAlaAsnIleThrSerAspGly 368
 QY 999 GCTGCG---CTACCACTGTCAGTTCCGGCGCGCTGCGGCGGCGAGCGCAGCGGCT--- 1052
 Db 369 ValIcIyValIleuGlnIyIyMetValArgSerIleArgIyProIyGlnAsnSerAlaProPhe 388
 QY 1052 ----- 1052
 Db 389 GlnIleuGlnMetPheAspIleValIcYserThrSerTrpAlaAsnThrAspIyIyCys 408
 QY 1053 -----GGAGTGTGCCCGCGGG-----CCGCAACTGCACCAACGCT----- 1085
 Db 409 GlnIleuProGlyIleuArgAspAspGlnSerCysAspIleuProIleuSerCysSerCys 428
 QY 1086 -----GCAAGGCGCTTGGCGCGCGGCGACACCGCTACT 1115
 Db 429 SerIleuThrAsnGlnValIcIyAlaIleuGlnProIleuGlyProIleuGlyProGlyIleuArg 448
 QY 1116 GGTGACCGTGCACCGCGCCTTCGCTGCGGCGCGGAGAGCGCGCTGTCGCAAGGCTG 1175
 Db 449 GlyProIyIyGlnGlnIcIyGlnIcIyGlnProGlyProIyIyGly----- 461
 QY 1176 CAGCCCGGAGCGCGCGCGCGCGCGC-----CCAGCGCCGCTGCGCGCGCGCGAC 1226
 Db 462 ----ProAspGlyProArgGlyIleuIleGlyIleuPro-GlyProGlnGlyProProGlyP 480
 QY 1227 CCGGGGAGCGCGCACCGCTGAGC 1249
 Db 480 roGlnGlyProSerGlyIleuSer 487
 RESULT 14
 US-09-252-991A-18035
 ; Sequence 18035, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 3142
 ; SEQ ID NO 18035
 ; LENGTH: 783
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-18035
 Alignment Scores:
 Pred. No.: 3,29e-10 Length: 783
 Score: 264.50 Matches: 154
 Percent Similarity: 33.8% Conservative: 21
 Best Local Similarity: 29.8% Mismatches: 176
 Query Match: 11.1% Gaps: 16
 DB: 2 Gaps: 31

US-10-699-035A-5 (1-1254) x US-09-252-991A-18035 (1-783)

QY 5 TCCTGAGAGCGCTGGCTGGCTG-----GAGCTTGGCGCTGG 46
:::|||||
Db 297 AAlaProGlyArgArg-ArgProProProAlaProSerArgThrProAlaProAlaG1 316
47 CGCTGGCGGAGCGGCGCGA-----GGCGGTCCACGACATCAG 88
|||
Db 316 YAlaAlaProGlyArgProGlyH1SerAlaProAlaProCysLeuArgAlaProAlaG1C1 336
|||
QY 89 CCCCCGAGGGGAGCGTGTCTGCTGAGCAGCTCAGCGCGTCTGACTAGAGT 148
|||||
Db 336 YProProArgThrProGly---ProProGlyAspAlaSerArgThrAlaSerGlyPr 355
|||||
QY 149 TCTCCCGGCTTGGGAGTTTGGGAGCTGGTGGCTCAGTCCCTGGGCGAGCGG 208
|||||
Db 355 Oh1SerProGlyAlaAlaProAlaGlnAlaArgInProArgAlaProAlaAlaArgAr 375
|||||
QY 209 CCTCGCTGCGAGTCTGTGCAAGTGGGAGTGGCCATACACCGAGTTCCCTTCGGCC 268
|||
Db 375 GPro-----ArgGlyThrGlnGlySerSerSerAlaProArgArgThrGlyAlaAs 392
|||
QY 269 AGCAGAGCTCGG---TGAGGCTGCCAGGATCGGTGGGTCTTGGCCGAGCGCA 322
|||||
Db 332 pAlaGlnArgGlyGlnProAlaG1YAspProGlyProAlaGlnHisProProArgAr 412
|||||
QY 323 TGGGAGACACCCAGCTGGCTGGCTGGTCTATGCCAA-----GGAACAGC 370
|||||
Db 412 GAla-----TTrpProGlyAspProArgCysArgProAspProAlaG1YThrAl 428
|||||
QY 371 TCTTTGCTAGACATAGAGTCCCGGCGAGGGTCCCAAGTCTGTGGGTGAGCAG 430
|||
Db 428 aPro-----GlnGlyAlaProArgGlyArgArgThrGlyAlaGly----- 442
|||
QY 431 ATGGGCGCTCCAGCGACCTGTGGGCGCCCATGCGAGA----- 470
|||
Db 443 ---ArgGlyGln-----ProProAspAlaG1YAspProValAlaAlaGlyLe 457
|||
QY 471 -----GCTCAGAGACCTGGG----- 485
|||||
Db 457 uSerArgAspAspProProGlnProThrGlyProGlyLeuProGlnArgGlyYAspProPr 477
|||||
QY 486 -----CGTACCGGTTCATTGTACGACCGGCGG----- 515
|||
Db 477 oArpPheSerAlaProArgArgProAlaProGlnHisArgProLeuG1YH1SerProAl 497
|||||
QY 516 ---AGGCAACTTCGAGAGCTGTCAGCGGCTGCCAGCCCTGCCGAGAGACCTGC 571
|||||
Db 497 aAlaArgGln---ProGlyGlyThrSerArg----- 506
|||||
QY 572 ACTTGTGAGCGTGAATGACCTGCACATATTGTCCAGAGCTGAGGGCTCCAT---TC 628
|||||
Db 507 -----ArgAlaProProAlaThrHis---ProAlaGlyArgGlyAlaTyThrG1 522
|||||
QY 629 TGGCGATGGCGCGGCGAGCTCCATGTCACGGAATCACTGCAGCGGCTTCGCGCTG 688
|||||
Db 522 YArgArgThrAlaG1YAla-----LeuProProAl 533
|||||
QY 689 CCGGCGACCCCGTGCAGCGAGACTCGGCGCTACTAGTGTGAGCTGAGCTGGCCGAGG 748
|||||
Db 533 a-----ThGlyGlnArgArgGlyArgAlaValAlaSerValGly----- 546
|||||
QY 749 CCCAGCGGAGGCTGCAGACGCCAGACAGCTGCCAGGAGAACGACGAGCTGAT----- 803
|||||
Db 547 -ProAlaGlyAlaAlaProGlyGlnGlyAlaProGlyGlyArgArgGlyLeuHisArgAr 566
|||||
QY 804 -----CTGGCGCGGCTTCGACCCGGA-- 824
|||||
Db 566 gProAspLeuProAlaAlaAlaGlyLeuHisProAlaLeuProAlaArgH1SerGlyAlaG1Y 586
|||||
QY 825 ---CACGGA-----CTACGAGCTGGCGCTATGCTGCTAGTCCAACTGGCGGCTCC 871
|||||

Db 586 YProHisGlyArgGlyGlnArgArgGlyAlaG1YGlnGlyArgArgArgPro-- 605
QY 872 TGAGGCCCCAGATCTCGGGGTGGCGACGCGGACAGAGAGCGCG---GCCAGCGCA 928
|||
Db 606 -----GlyProGlyLeuArgProG1YArgGlyYArgArgThrGlnArgAlaAr 621
|||||
QY 929 TGTGATCTCCACGCGCGCGCGGAGCTCCCGCTGAGTGGGAGTGGGAGCGGCTG 988
|||||
Db 621 gArgHisProProAlaProArgProAspHisProArgGlnProCysProArgArgThrPr 641
|||||
QY 989 CAGCGCGGAGCTCGGCTTACACGAGCTGAGTCCGGGCGCTCGGAG----- 1034
|||||
Db 641 oArgArgGlyAlaAsp---AlaArgAlaAlaGlnArgValAlaGlyValaAspProGlnG1 660
|||||
QY 1035 ---CGGAGAGCGGAGCGGTGAAGTGGCCGCGGCGGAGCTGACGACGCTGCAGG 1090
|||||
Db 660 nh1SerHisGlyProThrGlnGlyYAspArgArg---ProThrGlyAlaHisArgAlaTy 679
|||||
QY 1091 GCTTGGCGCGGAGCGGCTTACCTGTGACCGGTGACCGCGGCTTCGGCTCGGCGCGG 1150
|||||
Db 679 sAlaGlnAlaAlaHisGlnLeuGlyGlyYAlaAspGlnPhe-----Ar 694
|||||
QY 1151 AGAGCGCGCTTCCGCGAAGCGCTGCAC----- 1178
|||||
Db 694 gGlnValArgHisArgArgGlyLeuHisAlaGlyValaAspGlyInProArgArgArgG1 714
|||||
QY 1179 -----GCCGACGCGCGCGCGCGCC 1198
|||||
Db 714 YArgAlaTyPrProLeuProAlaAlaG1YProProG1YAlaAlaArgArgProArgAlaAs 734
|||||
QY 1199 GCCACGCGCGCGTCCGCGCGCGCGCGGAC---CCGCGGAGCGGCGCGCG 1244
|||||
Db 734 pArgGlnProGlnGlyProArgAlaAspArgGlyGlyYAspArgLeuPro 750
|||||

RESULT 15
US-09-252-991A-19127
; Sequence 19127, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19127
; LENGTH: 590
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19127

Alignment Scores:
Pred. No.: 3,666-10 Length: 590
Score: 263.50 Matches: 156
Percent Similarity: 32.84 Conservative: 24
Best Local Similarity: 28.54 Mismatches: 170
Query Match: 11.14 Indels: 198
DB: Gaps: 34

US-10-699-035A-5 (1-1254) x US-09-252-991A-19127 (1-590)

QY 1244 CGGCTGGGCGGTCCCGGGGTGGGCGCGGAGCGGCGTGGGCGCGGCGCGCGCG 1185
|||||
Db 78 ArgLeuAlaAlaProGlyGlyYAlaHis-----ArgGlyAlaTrpArgAspGln 94
|||||
QY 1184 TCGGCGTGCAGCGCTTGGCGGAGCAGCGCGCTGTCGCGCGCGGAGGAGCGCGCG-- 1127
|||||
Db 95 ProGlyAlaGln-ProGlnArgGlyAlaProGlyHisGlyProGlyGlnArgThrG1 114
|||||

```

QY 1126 -----TCACGGTACCCAGTAGAGCGGTGCCCCGGCCAGCGCTTCGACGCTGTGCAG 1074
DB 114 yValValGlyLeuGlyProGlyGluArg-----ProGlyProAlaPheLeuSerAr 131
QY 1073 TTGCGGG----- 1067
DB 131 gGlyGlyLeuGlnAraGlyArgGlyAraGlySerArgGlyLeuSerGlnAlaAspArgGlyHis 151
QY 1066 -CCGGCGGACCTCCACCCGCTGCGGCTCCGCGCCCGACGCGGCGCCGAATGCAGCTGG 1008
DB 151 sProProAraGlyProAlaAlaHisProSerArgPro-----ArgArgThiSerGlnI 169
QY 1007 TAGCCGAGCGCGCGCTGACGCCGACGCTGGGGCCCACTCACGCGAGCTGCGCGCG 948
DB 169 y-----AlaGlnAraGlyAlaValProGlyGlyValProAraGlyAraGlyProGly----- 184
QY 947 CGGGGCTGGGAGATAGACGATGCGCTGCGCGCGCTTCCTGCGCGCGCTGGGACCCCG 888
DB 185 -----GlyAraGlyLeuAlaLeuAlaGlyGlySerTyrProGlyHisGlySerArgProAl 202
QY 887 AGGATCTGGG-----GCCTCAGAGAGCGCACCTTGGACTCAGAGCACTAGCGGCAG 837
DB 202 aGlyThrGlyAraPheAlaAspAlaArgHisAlaGlnAraGlnHisGlnAraGlyAlaProGly 222
QY 836 TCGTAGTCCGTCGCGGTCGAGGC-----CGGCGCCAGATCCAGTCC 795
DB 222 yAraGlnAlaAlaIleThrGlyAraGlyAspSerValArgGlyGlyGlnAraGlyHisAlaGlyAspTr 242
QY 794 GTGGGCTTCCTGGGACGCTGCTGGC----- 770
DB 242 aArgSerGlyLeuGlnProAlaGlyGlyGlnProGlyLeuGlnSerHisHisArgLeuSe 262
QY 769 -----GTCCTGACGCCCGCGGCTGGGCGCTGG-----GC 741
DB 262 rProAraGlyAraGlyAraGlyAlaGlnAraGlnProAlaGlyGlyTyrProAraGlyProAl 282
QY 740 ACCAGCTCCAGCA-----CATAGTACCCGAGCTGCGGTCAACAGGGGTGGCCAG 690
DB 282 aProLeuProAlaAspProAraGlyAraGlyAlaProAlaArgAlaValAlaGly-----Ar 300
QY 689 GCCAGCGCGA-----AGCCGCTGG---ACGTGATTCCTGTGCGATGAGCTGCTCGGC 639
DB 300 gThrGlyGlyTyrAlaGlnGlyIleThrArgSerLeuSerProMetAlaAlaAlaGlnProAl 320
QY 638 CGCATCGCGAGATGAGCCCTCAGCTCTTGACAATGATGTCAGTCA----- 587
DB 320 aAlaAraGlyGlnTyrAraGlyAspPro-----HisArgArgPh 333
QY 586 -----CCAGTCCACAAAGTCAGAGTGTCTTCGCGAGGGGCT 549
DB 333 eleuHisArgSerGlnCysProProAraGlyAraGlyAlaProAlaSerProAlaAlaIle 353
QY 548 GAGGACGCGCTGACAGCTCCAGAAATTGCTCGGCGCGCTGTCAATGAACAGGTG 489
DB 353 uArg-----ArgIleAspArgProGly----- 360
QY 488 AGGCCCAAGT-----CCTTAGCTTCCTGATGGGGGGCCCAAGGGTCCGTGAGCGG 435
DB 361 -GlnProProValValProAraGlyAlaAlaAraGlyGlyProAlaGly----- 376
QY 434 CCATGTCACCCACACACGACTTTGGGCAACCCCTGGCGGCACTGATGCTTCAGCA 375
DB 377 -----AlaProGlyGlyGlyAraGlyAlaGlyProAlaAla----- 387
QY 374 AACAGCTGTCCTTGCGCATAGACAGCGCCAGGCAAGTGT----- 335
DB 388 -----AspArgProGlyProLeuGlnAlaAlaGlnAraGlnProAraGPr 402
QY 334 -----GGGTGTCACCCATGGCTGGGGAGAGACGCGACCGCATCCT----- 293
DB 402 oAraGlyAraGlyAraGlyProValAla-----GlnIleAspGlyProAlaProAlaAlaHisAraGAl 421

```

```

QY 292 -GGGACGCTCACCCGAGCTGTGCTGGCCGAAGGGAACTGGGTATGCGCGACTGCC 234
DB 421 aGlySerGlnHisProGlyAlaSerGlyAraGlyAraGlyValArg----- 435
QY 223 AGGTACACAGACTGG-----CAGCGAGGCGCCGGTGCACGAGGGGAGTGAAGCC 183
DB 426 -HisSerProAraGlyTyrAraGlnHisAraGlyAraGlyProValAlaProGly-----ArgTh 453
QY 182 ACCAGTGCACCCCAAACTCC-----GAACCCGGGAGAACTGTAGTGAAGACGCTG 129
DB 453 rProAlaGlyProValAlaProAlaGlyGlnAraGlyAraGlyProAraGlyAraProAraGAr 473
QY 128 GCTGAGC-----TGTCCAGCAGAAATCAGT----- 101
DB 473 gLeuAlaGlyAraGlnProValPheAraGlyAlaGlyThrGlyAraLeuAraGlyAraGlyAra 493
QY 100 -----CCCTCGGGGG 90
DB 493 oGlyGlyProGlyAraAlaThrCysGlnAlaThrAraGlyGlnTyrLeuProValLeuPr 513
QY 89 GCTGATGCTGTGACCGCGCTCCGCGCGCTCCGCGCCAGCG---CCAGCCGAAGCTC 33
DB 513 oAraGlnProAlaGlyGlnHisProAraGlyThrProAlaProGlyAraProValAlaGlnI 533
QY 32 AGGG-----CCAGCGGAGCGCC 15
DB 533 yGlyAraGlyGlyProAlaGlyAraGlyAraLeuProAlaGlnAlaValProGlyGlnAraGPr 553
QY 14 G-----TCCAGGGAGCA 2
DB 553 oProGlyLeuGlyAraGlyAla 560

```

Search completed: February 13, 2006, 13:44:00
Job time : 72.2111 secs

THIS PAGE BLANK (USPTO)